

计量经济学

第七讲

异 方 差 性

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) (0.644284)
$$t = (-1.931062) (8.340265)$$
 $R^2 = 0.785456 \quad \bar{R}^2 = 0.774146 \quad F = 69.56003$

式中 γ 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人。

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?



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



goromety. 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,...,n)有:

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

因为方差是度量被解释变量Y的观测值围绕回归线

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



是 是 是 是 是 性的含义

设模型为

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

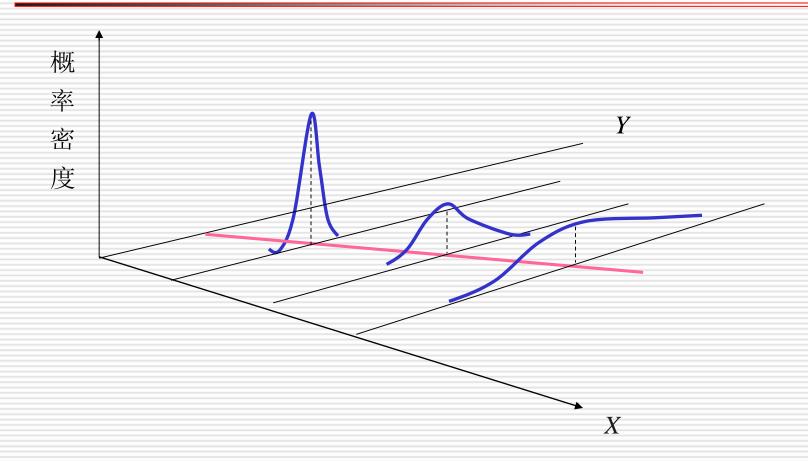
如果对于模型中随机误差项 u_i 有:

$$Var(u_i) = \sigma_i^2$$
, $i = 1, 2, 3, ..., n$ $(5, 3)$ 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

$$Var(u_i) = \sigma_i^2 = \sigma^2 f(X_i)$$
 (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

$$Y_{i} = \beta_{1} + \beta_{2} X_{2i} + u_{i}^{*}$$
 (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) 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。



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



一、图示检验法

(一) 相关图形分析

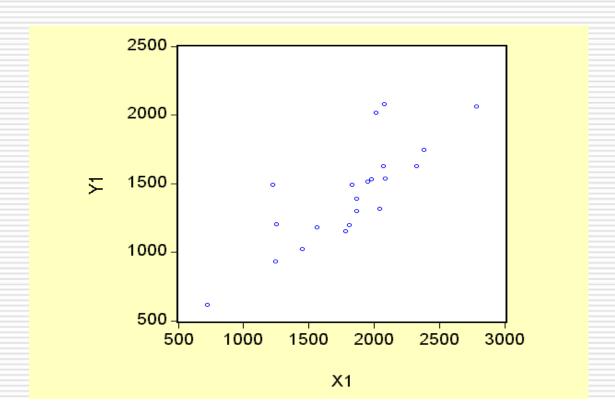
方差描述的是随机变量取值的(与其均值的)离散程度。因为被解释变量 γ 与随机误差项u有相同的方差,所以利用分析Y与X的相关图形,可以初略地看到Y的离散程度与X之间是否有相关关系。

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:

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

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。



上のできた。 ニ、Goldfeld-Quanadt松絵

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。



岛 (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

$$H_0: \sigma_i^2 = \sigma^2, i = 1, 2, ..., n;$$
 $H_1: \sigma_1^2 \le \sigma_2^2 \le ... \le \sigma_n^2$



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 $\sum e_{i}^{2}$ 和 $\sum e_{2i}^{2}$ 。 $\sum e_{i}^{2}$ 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:

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

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

其中 v_t为随机误差项。



$1. 求回归估计式并计算 <math>e_t^2$

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

$$\hat{e}_{t}^{2} = \hat{\alpha}_{1} + \hat{\alpha}_{2}X_{2t} + \hat{\alpha}_{3}X_{3t} + \hat{\alpha}_{4}X_{2t}^{2} + \hat{\alpha}_{5}X_{3t}^{2} + \hat{\alpha}_{6}X_{2t}X_{3t}$$
 (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

$$H_0: \alpha_2 = ... = \alpha_6 = 0$$
, $H_1: \alpha_j (j=2, 3, ..., 6)$ 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 $\chi^2_{\alpha}(5)$,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。

第四节 异方差性的补救措施

主要方法:

- ●模型变换法
- 加权最小二乘法
- 模型的对数变换



一、模型变换法

以一元线性回归模型为例:

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

经检验 u_i 存在异方差,且

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

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:

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

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

则有:

$$Y_{i}^{*} = \beta_{1}^{*} + \beta_{2} X_{i}^{*} + \nu_{i}$$



随机误差项 v_i 的方差为

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

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	$var(u_i)$	${oldsymbol{\mathcal{V}}}_i$	$var(v_i)$
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



S二、加权最小二乘法

以一元线性回归模型为例:

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

经检验 u_i 存在异方差,且:

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

其中 σ^2 是常数, 是 X_i 的某种函数。



(一) 基本思路

区别对待不同的 σ_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 $\sum e_i^2$ 更好地反映 σ_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 $w_i = 1/\sigma_i^2 (i = 1, 2, ..., n)$, 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:

$$\sum w_i e_i^2 = \sum w_i (Y_i - \beta_1^* - \beta_2^* X_i)^2$$



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:

$$\hat{\beta}_{1}^{*} = \bar{Y}^{*} - \hat{\beta}_{2}^{*} \bar{X}^{*}$$

$$\hat{\beta}_{2}^{*} = \frac{\sum w_{i}(X_{i} - \bar{X}^{*})(Y_{i} - \bar{Y}^{*})}{\sum w_{i}(X_{i} - \bar{X}^{*})^{2}}$$

其中:

$$\overline{X}^* = \frac{\sum w_i X_i}{\sum w_i}, \ \overline{Y}^* = \frac{\sum w_i Y_i}{\sum w_i}$$



三、模型的对数变换

在经济意义成立的情况下,如果对模型:

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

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。



岛的 第五节 案例分析

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表示卫生医疗机构数,X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			

CONOME EX.

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 X	-563.0548 5.373498	291.5778 0.644284	-1.931062 8.340265	0.0685 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood	0.785456 0.774164 623.0330 7375233. -163.8734	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion F-statistic Prob(F-statistic)		1588.238 1311.037 15.79747 15.89695 69.56003

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 $y \in X$,点 "ok",得样本回归估计结果,见教材表5.2。

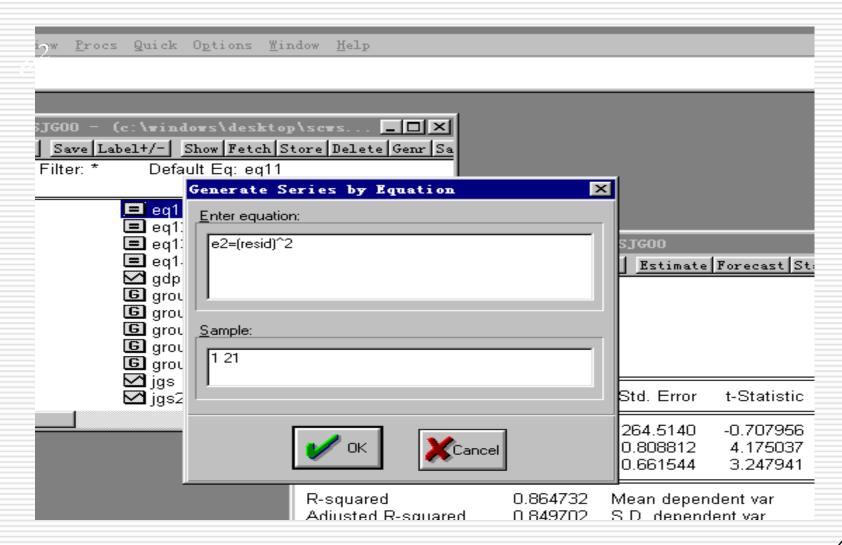


(1) 生成残差平方序列。

在得到表5.2估计结果后,用生成命令生成序列,记为 e^2 。 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: $e^2 = \operatorname{resid}^2 2$



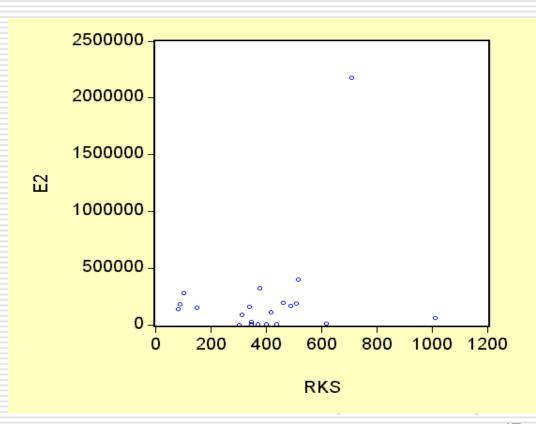
生成序列图示





(2) 绘制 e_t^2 对 X_t 的散点图。选择变量名 X 与 e^2 。(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 "Ascenging", 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,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: 18

Included observations: 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C X	598.2525 1.177650	119.2922 0.490187	5.015018 2.402452	0.0024 0.0531
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-VVatson stat	0,490306 0,405357 155,4343 144958,9 -50,57056 1,656269	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion F-statistic Prob(F-statistic)		852.6250 201.5667 13.14264 13.16250 5.771775 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|rrocs|UDjects| rrint|Mame|rreeze| rstimate|rorecast|Stats|nesids|

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 Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	0.966949	Mean dependent var		2520.750
	0.981441	S.D. dependent var		1781.608
	349.8466	Akaike info criterion		14.76518
	734355.8	Schwarz criterion		14.78504
	-57.06074	F-statistic		175.5379
	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。



表5.5

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

F-statistic		Probability	0.000000
Obs*R-squared		Probability	0.000119
One K-edualen	10.00330	Flobability	0.000113

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

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

Sample: 121

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

$$\chi^2_{0.05}(2) = 5.9915$$

因为

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

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:

$$w1 = 1/X$$
 $w2 = 1/X ^2$ $w3 = 1/sqrt(X)$

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",

option,在对话框中点 weighted LS,在weighted 中输入" w^2 "再点ok,即出现加权最小二乘结果。



表 5.7

View|Procs|Ubjects| Print|Name|Freeze| Estimate|Forecast|Stats|Kesids|

Dependent Variable: Y Method: Least Squares

Date: 07/09/05 Time: 11:24

Sample: 1 21

Included observations: 21 Weighting series: W2

weighting series. vvz						
Variable	Coefficient	Std. Error	t-Statistic	Prob.		
C X	368,6090 2,952958	84.16870 0.822688	4.379407 3.589402	0.0003 0.0020		
Weighted Statistics						
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	0.938665 0.935437 276.0493 1447861. -146.7790 1.705980	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion F-statistic Prob(F-statistic)		808.6991 1086.410 14.16943 14.26891 12.88381 0.001955		
Unweighted Statistics						
R-squared 0.625222 Adjusted R-squared 0.605497 S.E. of regression 823.4555 Durbin-Watson stat 0.380523		Mean dependent var S.D. dependent var Sum squared resid		1588.238 1311.037 12883501		

估计结果:

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

。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

- 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



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。



