

§ 2.6 异方差性

Heteroskedasticity

- 一、异方差性的概念
- 二、异方差性的后果
- 三、异方差性的检验
- 四、异方差性的估计
- 五、案例

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。
- 如果随机误差项序列不具有同方差性，即出现**异方差性**。

一、异方差的概念

1、异方差的概念

对于模型

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

同方差性假设为 $Var(\mu_i) = \sigma^2 \quad i=1,2,\cdots,n$

如果出现

$$Var(\mu_i) = \sigma_i^2 \quad i=1,2,\cdots,n$$

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

2、异方差的类型

- **同方差性**假定的意义是指每个 μ_i 围绕其零平均值的变差，并不随解释变量X的变化而变化，不论解释变量观测值是大还是小，每个 μ_i 的方差保持相同，即

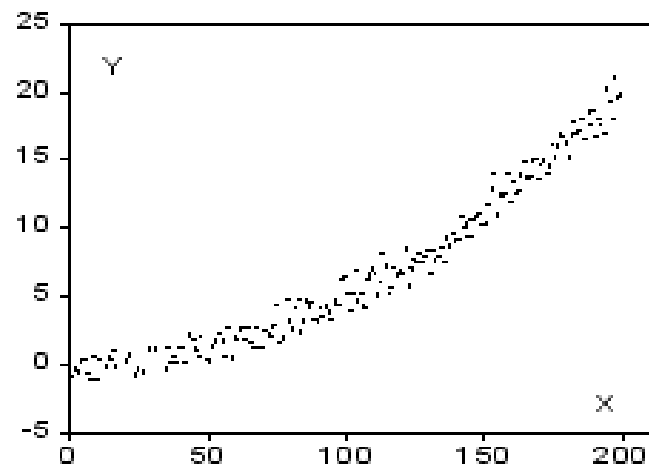
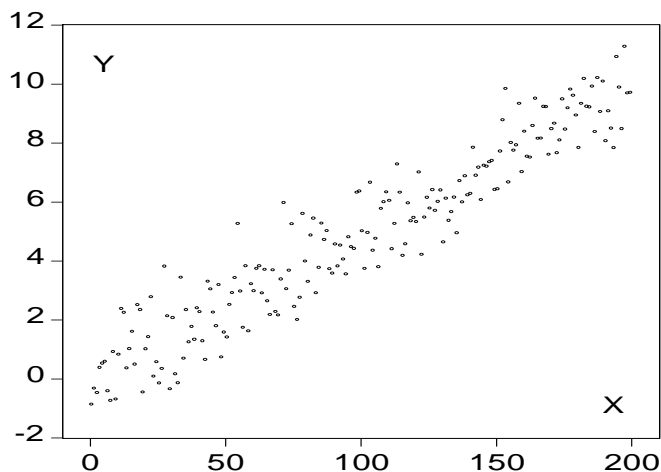
$$\sigma_i^2 = \text{常数}$$

- 在**异方差**的情况下， σ_i^2 已不是常数，它随X的变化而变化，即

$$\sigma_i^2 = f(X_i)$$

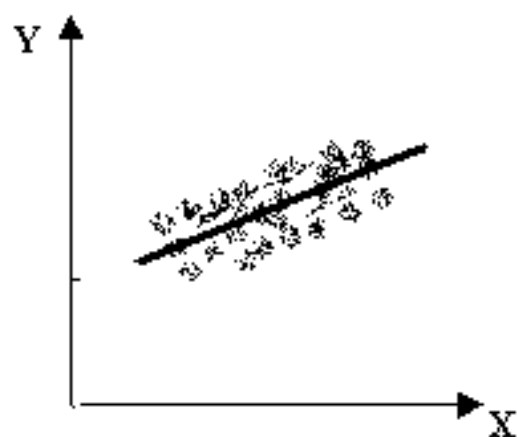
同方差假定：模型的假定条件(1) 给出 $\mathbf{Var}(u)$ 是一个对角矩阵，且主对角线上的元素都是常数且相等。

$$\mathbf{Var}(u) = \mathbf{E}(u u') = \sigma^2 \mathbf{I} = \sigma^2 \begin{bmatrix} 1 & \cdots & 0 \\ & 1 & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & & 1 \end{bmatrix}$$

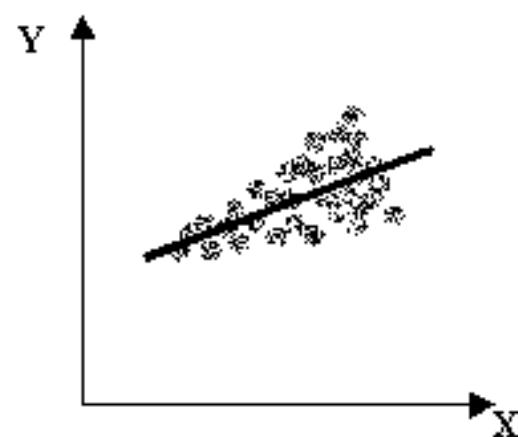


- 异方差一般可归结为三种类型：

- (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。



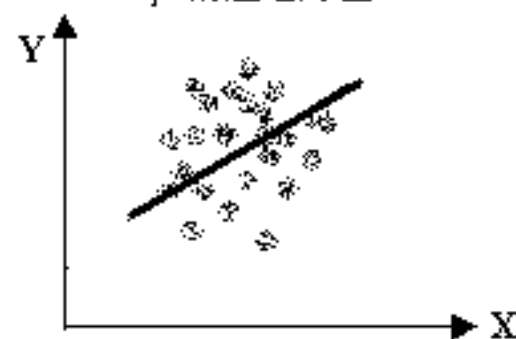
同方差



μ 的递增方差



μ 递减方差



复杂型

异方差概念

当这个假定不成立时， $\text{Var}(u)$ 不再是一个纯量对角矩阵。

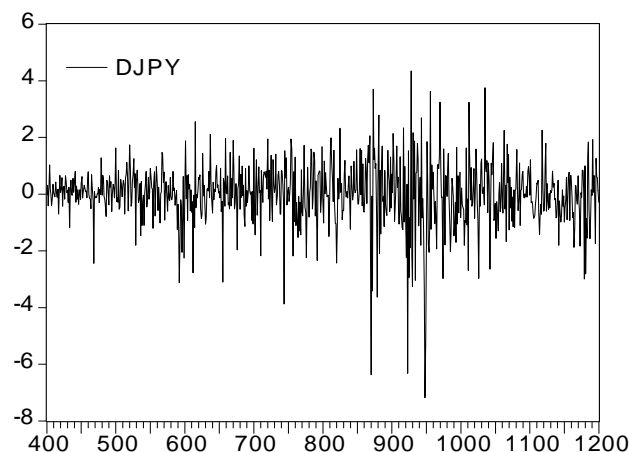
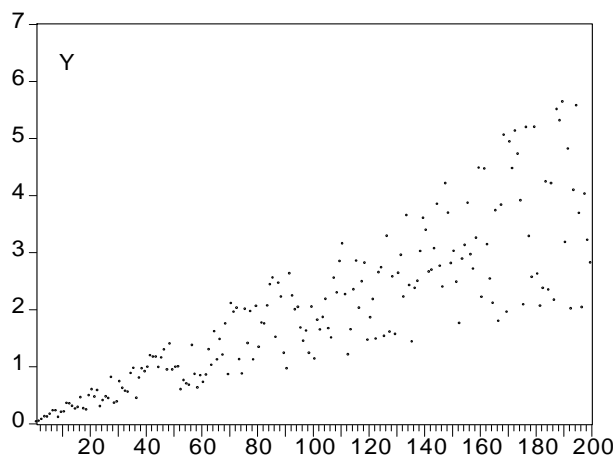
$$\text{Var}(u) = \sigma^2 \Omega = \sigma^2 \begin{bmatrix} \sigma_{11} & \cdots & 0 \\ & \sigma_{22} & \vdots \\ & & \ddots \\ 0 & \cdots & \sigma_{TT} \end{bmatrix} \neq \sigma^2 I$$

(第2版教材第112页)

(第3版教材第91页)

当误差向量 u 的方差协方差矩阵主对角线上的元素不相等时，称该随机误差系列存在异方差。非主对角线上的元素表示误差项之间的协方差值。若 Ω 非主对角线上的部分或全部元素都不为零，误差项就是自相关的。

异方差通常还表现为条件自回归型（金融市场，石油价格市场，期货市场等）。



3、实际经济问题中的异方差性

- 例如：在截面资料下研究居民家庭的储蓄形为

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

Y_i 和 X_i 分别为第 i 个家庭的储蓄额和可支配收入。

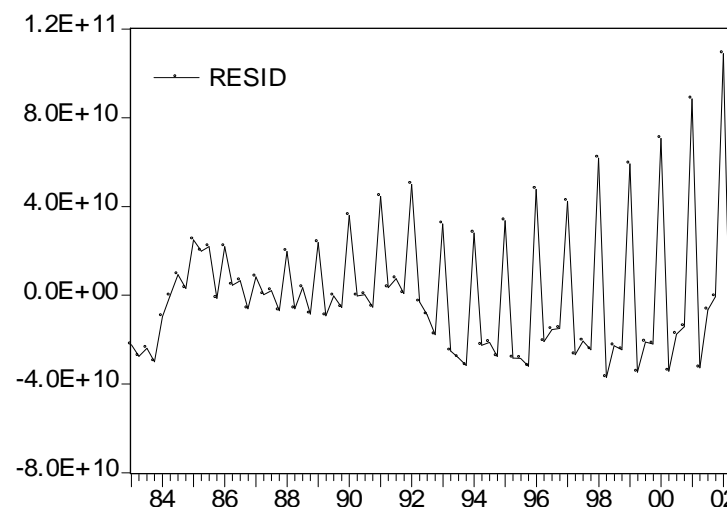
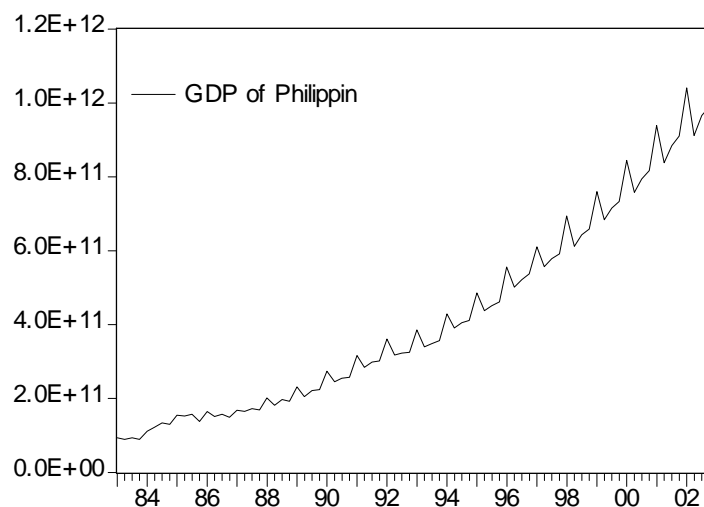
在该模型中， μ_i 的同方差假定往往不符合实际情况。对高收入家庭来说，储蓄的差异较大；低收入家庭的储蓄则更有规律性（如为某一特定目的而储蓄），差异较小。

因此， μ_i 的方差往往随 X_i 的增加而增加，呈单调递增型变化。

异方差来源

异方差来源:

- (1)** 时间序列数据和截面数据中都有可能存在异方差。
- (2)** 经济时间序列中的异方差常为递增型异方差。金融时间序列中的异方差常表现为自回归条件异方差。



- 例如，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：

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

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，处于中等收入组中的人数最多，处于两端收入组中的人数最少。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增大而先减后增。

如果样本观测值的观测误差构成随机误差项的主要部分，那么对于不同的样本点，随机误差项的方差随着解释变量观测值的增大而先减后增，出现了异方差性。

•例如，以某一行业的企业为样本建立企业生产函数模型

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

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。

二、异方差性的后果

1、参数估计量非有效

- 普通最小二乘法参数估计量**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

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

- 而且，在大样本情况下，参数估计量**仍然不具有渐近有效性**，这就是说参数估计量不具有**一致性**。

以一元线性回归模型为例进行说明：

(1) 仍存在无偏性：证明过程与方差无关

由于
$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \quad (2.4.1)$$

的参数 β_1 的 OLS 估计量 $\hat{\beta}_1$ 为：

$$\hat{\beta}_1 = \sum k_i Y_i = \beta_1 + \sum k_i \mu_i = \beta_1 + \sum \frac{x_i}{\sum x_i^2} \mu_i$$

故
$$E(\hat{\beta}_1) = E(\beta_1) + \sum \frac{x_i}{\sum x_i^2} E(\mu_i) = \beta_1 \quad (2.4.2)$$

(2) 不具备最小方差性

由于

$$\begin{aligned}\text{var}(\hat{\beta}_1) &= E(\hat{\beta}_1 - \beta_1)^2 = E\left(\sum \frac{x_i}{\sum x_i^2} \mu_i\right)^2 = \frac{E\left(\sum x_i \mu_i\right)^2}{\left(\sum x_i^2\right)^2} \\ &= \frac{\sum x_i^2 E(\mu_i^2)}{\left(\sum x_i^2\right)^2}\end{aligned}$$

(注：交叉项 $\sum_{\substack{i,j \\ i \neq j}} (x_i \mu_i)(x_j \mu_j)$ 的期望为零)

在 μ_i 为同方差的假定下，

$$\begin{aligned}\text{var}(\mu_i) &= E(\mu_i)^2 = \sigma^2 \\ \text{var}(\hat{\beta}_1) &= \frac{\sum x_i^2 \sigma^2}{\left(\sum x_i^2\right)^2} = \frac{\sigma^2}{\sum x_i^2}\end{aligned}\tag{2.4.3}$$

在 μ_i 存在异方差的情况下

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

假设 $f(X_i) = X_i^2$ ，并且记异方差情况下 β_1 的 OLS 估计为 $\tilde{\beta}_1$ ，则

$$\text{var}(\tilde{\beta}_1) = \frac{\sum x_i^2 \sigma^2 f(X_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2} \cdot \frac{\sum x_i^2 X_i^2}{\sum x_i^2} \quad (2.4.4)$$

对大多数经济资料有： $\sum x_i^2 X_i^2 / \sum x_i^2 > 1$ ，

比较(2.4.3)与(2.4.4)，

$$\text{var}(\tilde{\beta}_1) > \text{var}(\hat{\beta}_1) \quad (2.4.5)$$

2、变量的显著性检验失去意义

关于变量的显著性检验中，构造了 t 统计量

$$t = \hat{\beta}_i / S(\hat{\beta}_i) \quad (2.4.6)$$

在该统计量中包含有随机误差项共同的方差，并且有 t 统计量服从自由度为 $(n-k-1)$ 的 t 分布。如果出现了异方差性， t 检验就失去意义。

其它检验也类似。

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随机误差项共同的方差 σ^2 。

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。

三、异方差性的检验

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:

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

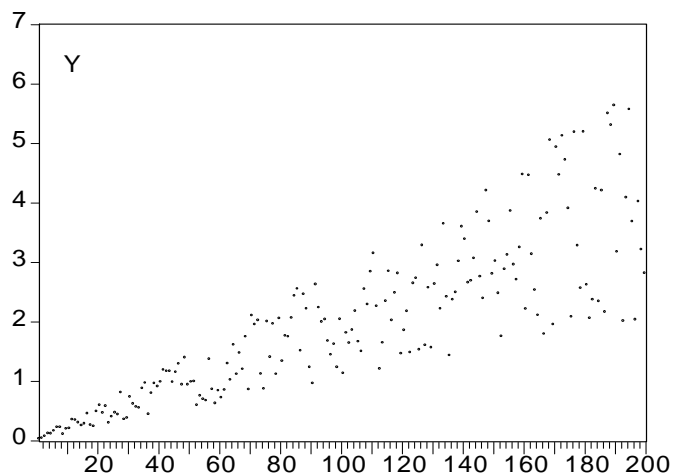
$$\begin{aligned}\tilde{e}_i &= Y_i - (\hat{Y}_i)_{OLS} \\ \text{Var}(\mu_i) &= E(\mu_i^2) \approx \tilde{e}_i^2\end{aligned}\tag{2.4.7}$$

即用 \tilde{e}_i^2 来表示随机误差项的方差。

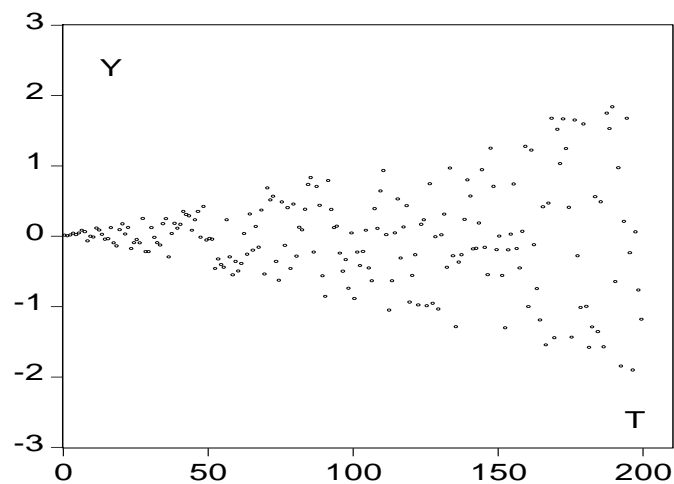
异方差检验

定性分析异方差

- (1) 宏观经济变量容易出现异方差。
- (2) 利用散点图做初步判断。
- (3) 利用残差图做初步判断。



散点图

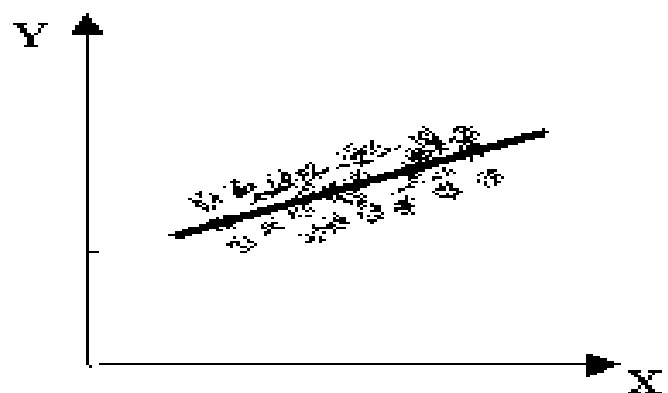


残差图

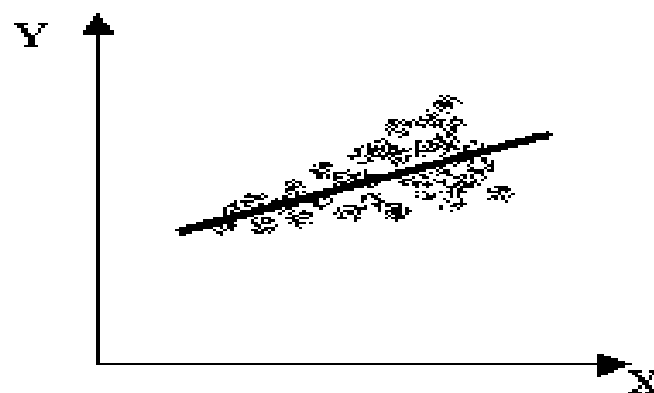
2、图示检验法

(1) 用X-Y的散点图进行判断

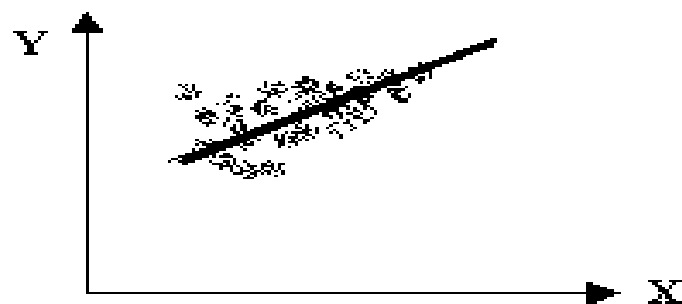
看是否存在明显的散点扩大、缩小或复杂型趋势（即不在一个固定的带型域中）



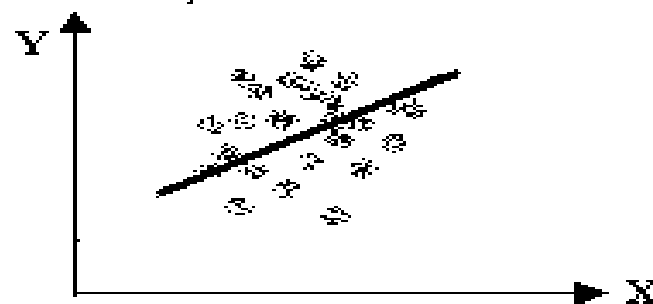
同方差



μ 的递增方差



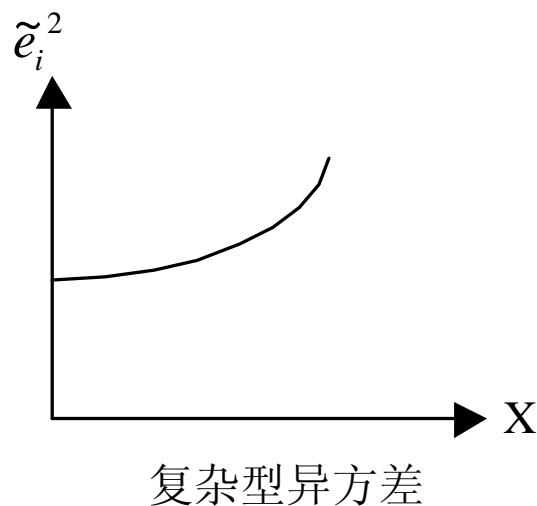
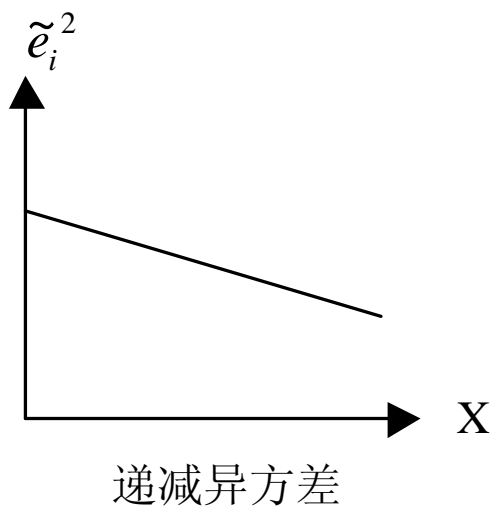
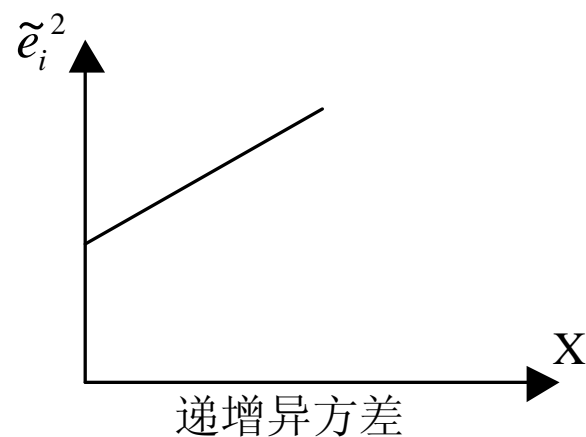
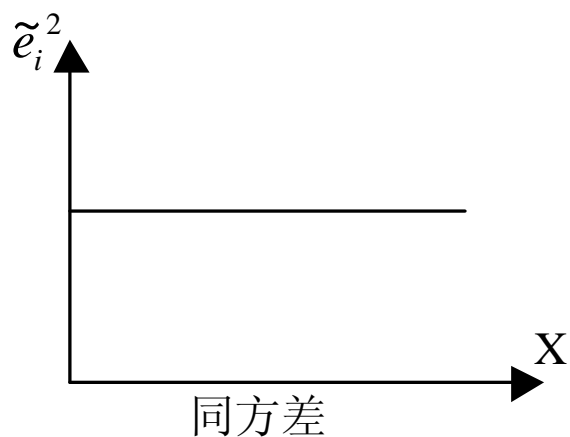
μ 递减方差



复杂型

(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



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) 按解释变量观察值 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对每个子样分别求回归方程，并计算各自的残差平方和。

分别用 $\sum \tilde{e}_{1i}^2$ 与 $\sum \tilde{e}_{2i}^2$ 表示对应较小 X_i 与较大 X_i 的子样本的残差平方和(自由度均为 $\frac{n-c}{2} - k - 1$)

④提出假设: $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$

σ_1^2 与 σ_2^2 分别为两个子样对应的随机项方差。

⑤构造统计量

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

⑥检验。给定显著性水平 α ，确定 F 分布表中相应的临界值 $F_\alpha(v_1, v_2)$ 。

若 $F > F_\alpha(v_1, v_2)$ ，存在递增异方差；

反之，不存在异方差。

(2) 戈里瑟 (Gleiser) 检验与帕克 (Park) 检验

• 戈里瑟检验与帕克检验的思想:

以 $|\tilde{e}|$ 或 \tilde{e}_i^2 为被解释变量, 以原模型的某一解释变量 x_j 为解释变量, 建立如下方程:

$$|\tilde{e}_i| = f(X_{ji}) + \varepsilon_i \quad i=1,2,\cdots,n \quad (\text{Gleiser})$$

或

$$\tilde{e}_i^2 = f(X_{ji}) + \varepsilon_i \quad i=1,2,\cdots,n \quad (\text{Park})$$

选择关于变量 x_j 的不同的函数形式 (如 $f(X_{ji}) = X_{ji}^2$ 或 $f(X_{ji}) = \sigma^2 X_{ji}^\alpha e^{v_i}$), 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。

异方差检验

Glejser检验（直接拟合法）

检验 $|\hat{u}_t|$ 是否与解释变量 x_t 存在函数关系。若有，则说明存在异方差；若无，则说明不存在异方差。通常应检验的几种形式是

$$|\hat{u}_t| = a_0 + a_1 x_t \quad |\hat{u}_t| = a_0 + a_1 x_t^2 \quad |\hat{u}_t| = a_0 + a_1 \sqrt{x_t}, \dots$$

Glejser 检验的特点是：

- ①既可检验递增型异方差，也可检验递减型异方差。
- ②一旦发现异方差，同时也就发现了异方差的具体表现形式。
- ③当原模型含有多个解释变量值时，可以把 $|\hat{u}_t|$ 拟合成多变量回归形式。

（4）自回归条件异方差（ARCH）检验（不要求掌握）

（5）Spearman 等级相关系数检验（不要求掌握）

如 Park 检验法中，对一般的方程形式：

$$f(X_{ji}) = \sigma^2 X_{ji}^\alpha e^{v_i}$$

通过

$$\ln(\tilde{e}_i^2) = \ln \sigma^2 + \alpha \ln X_{ji} + v_i$$

检验 α 的显著性，若存在统计上的显著性，表明存在异方差性。

注意：

由于 $f(\mathbf{X}_j)$ 的具体形式未知，因此需要进行各种形式的试验。

异方差检验

(3) White检验

White检验由H. White 1980年提出。White检验不需要对观测值排序，也不依赖于随机误差项服从正态分布，它是通过一个辅助回归式构造 χ^2 统计量进行异方差检验。以二元回归模型为例，White检验的具体步骤如下。

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t$$

①首先对上式进行OLS回归，求残差 u_t 。

②做如下辅助回归式，

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 x_{t1} + \alpha_2 x_{t2} + \alpha_3 x_{t1}^2 + \alpha_4 x_{t2}^2 + \alpha_5 x_{t1} x_{t2} + v_t$$

即用 \hat{u}_t 对原回归式中的各解释变量、解释变量的平方项、交叉积项进行OLS回归。注意，上式中要保留常数项。求辅助回归式的可决系数 R^2 。

③White检验的零假设和备择假设是

H_0 : u_t 不存在异方差, H_1 : u_t 存在异方差。

(3) White检验

④在同方差假设条件下，统计量

$$TR^2 \sim \chi^2(5)$$

其中 T 表示样本容量， R^2 是辅助回归式的OLS估计的可决系数。自由度5表示辅助回归式中解释变量项数（注意，不计算常数项）。 TR^2 属于LM统计量。

⑤判别规则是

若 $TR^2 \leq \chi^2_{\alpha}(5)$, 接受 H_0 (u_t 具有同方差)

若 $TR^2 > \chi^2_{\alpha}(5)$, 拒绝 H_0 (u_t 具有异方差)

四、异方差性的估计

——加权最小二乘法(WLS)

Weighted Least Squares

1、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，然后采用普通最小二乘法估计其参数。
- 例如，在递增异方差下，对来自较小 X_i 的子样本，其真实的总体方差较小， Y_i 与回归线拟合值之间的残差 e_i 的信度较大，应予以重视；而对较大 X_i 的子样本，由于真实总体的方差较大，残差反映的信息应打折扣。

- 加权最小二乘法就是对加了权重的残差平方和
实施OLS法：

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。

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

2、一个例子

- 例如，如果在检验过程中已经知道：

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

即随机误差项的方差与解释变量 x_j 之间存在相关性，那么可以用 $\sqrt{f(x_j)}$ 去除原模型，使之变成如下形式的新模型：

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

在该模型中，存在

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

即满足同方差性。于是可以用 **OLS** 估计其参数，得到关于参数 $\beta_0, \beta_1, \dots, \beta_k$ 的无偏的、有效的估计量。这就是加权最小二乘法，在这里权就是 $1/\sqrt{f(X_{ji})}$ 。

3、一般情况

对于模型

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{N} \quad (2.4.8)$$

存在

$$E(\mathbf{N}) = 0$$

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

$$\mathbf{W} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix} \quad (2.4.9)$$

即存在异方差性。

设

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

其中

$$D = \begin{pmatrix} \sqrt{w_1} & & \\ & \ddots & \\ & & \sqrt{w_n} \end{pmatrix}$$

用 \mathbf{D}^{-1} 左乘(2.4.8)两边, 得到一个新的模型:

$$\mathbf{D}^{-1}\mathbf{Y} = \mathbf{D}^{-1}\mathbf{X}\mathbf{B} + \mathbf{D}^{-1}\mathbf{N} \quad (2.4.10)$$

即 $\mathbf{Y}^* = \mathbf{X}^*\mathbf{B} + \mathbf{N}^*$

该模型具有同方差性。因为

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

于是，可以用 OLS 法估计模型(2.4.10)，得

$$\begin{aligned}\hat{\mathbf{B}} &= (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{Y}^* \\ &= (\mathbf{X}' \mathbf{D}^{-1'} \mathbf{D}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{D}^{-1'} \mathbf{D}^{-1} \mathbf{Y} \\ &= (\mathbf{X}' \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{-1} \mathbf{Y}\end{aligned}\tag{2.4.11}$$

这就是原模型(2.4.8)的加权最小二乘估计量，它是无偏、有效的。

这里权矩阵为 \mathbf{D}^{-1} ，它来自于矩阵 \mathbf{W} 。

4、求得权矩阵W的一种实用方法

从前面的推导过程看，它来自于原模型(2.4.8)残差项N的方差-协方差矩阵，因此仍然可对原模型(2.4.8)首先采用OLS法，得到随机误差项的近似估计量，以此构成权矩阵的估计量，即

$$\hat{\mathbf{W}} = \begin{bmatrix} \tilde{e}_1^2 & & & \\ & \tilde{e}_2^2 & & \\ & & \ddots & \\ & & & \tilde{e}_n^2 \end{bmatrix}, \quad \mathbf{D}^{-1} = \begin{bmatrix} 1/|\tilde{e}_1| & & & \\ & 1/|\tilde{e}_2| & & \\ & & \ddots & \\ & & & 1/|\tilde{e}_n| \end{bmatrix}$$

(2.4.12)

5、加权最小二乘法具体步骤

①选择普通最小二乘法估计原模型，得到随机误差项的近似估计量 \tilde{e}_i ；

② 建立 $1/|\tilde{e}_i|$ 的数据序列；

③ 选择加权最小二乘法，以 $1/|\tilde{e}_i|$ 序列作为权，进行估计得到参数估计量。

实际上是以 $1/|\tilde{e}_i|$ 乘原模型的两边，得到一个新模型，采用普通最小二乘法估计新模型。

异方差的修正方法

利用**Glejser**检验结果消除异方差

假设 **Glejser** 检验结果是

$$|\hat{u}_t| = \hat{a}_0 + \hat{a}_1 \mathbf{x}_t$$

说明异方差形式是 $\text{Var}(u_t) = (\hat{a}_0 + \hat{a}_1 \mathbf{x}_t)^2 \sigma^2$ 。用 $(\hat{a}_0 + \hat{a}_1 \mathbf{x}_t)$ 除原模型各项，

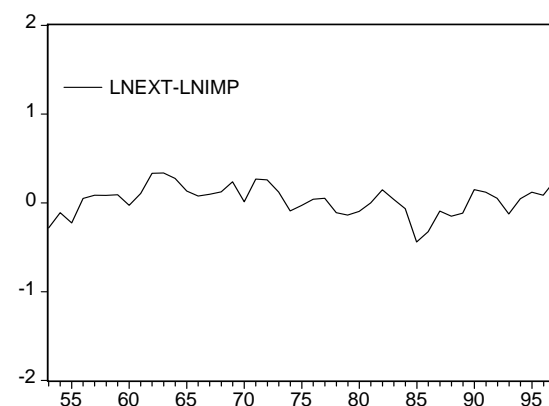
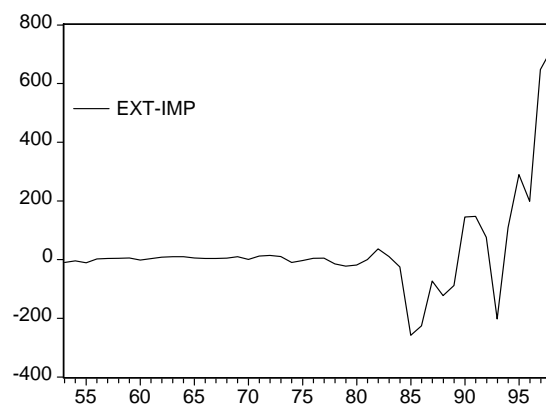
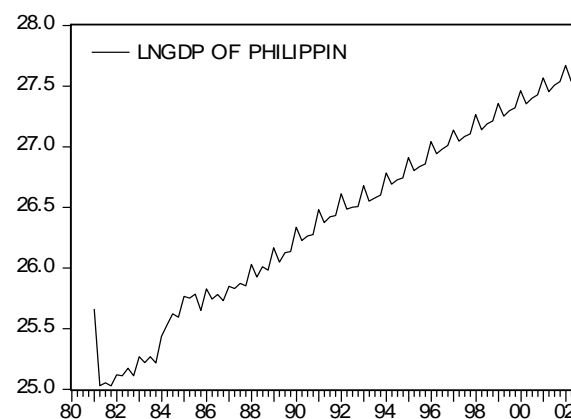
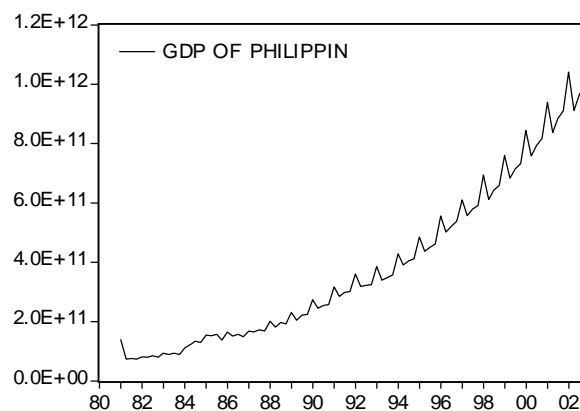
$$\frac{y_t}{\hat{a}_0 + \hat{a}_1 x_t} = \beta_0 \frac{1}{\hat{a}_0 + \hat{a}_1 x_t} + \beta_1 \frac{x_t}{\hat{a}_0 + \hat{a}_1 x_t} + \frac{u_t}{\hat{a}_0 + \hat{a}_1 x_t}$$

$$\text{则 } \text{Var}\left(\frac{u_t}{\hat{a}_0 + \hat{a}_1 x_t}\right) = \frac{1}{(\hat{a}_0 + \hat{a}_1 x_t)^2} \text{Var}(u_t) = \frac{1}{(\hat{a}_0 + \hat{a}_1 x_t)^2} (\hat{a}_0 + \hat{a}_1 x_t)^2 \sigma^2 = \sigma^2$$

消除了异方差。**OLS** 估计后，把回归参数的估计值代入原模型。

异方差的修正方法

通过对数据取对数消除异方差



中国进出口贸易额差（1953-1998）

对数的中国进出口贸易额之差

6、注意

- 在实际建模过程中，尤其是截面数据作样本时，人们通常并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。

- 在应用软件中，给出了权矩阵的多种选择。
例如在Eviews中给出了权矩阵的3种选择：**White权矩阵、Newey-West权矩阵和自己输入权矩阵。**
- **White, 1980, A heteroskedasticity-consistent covariance matrix and direct test for heteroskedasticity, Econometrica, 48, 817-38**
- **Newey, West, 1987, A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703-8**

五、案例—1

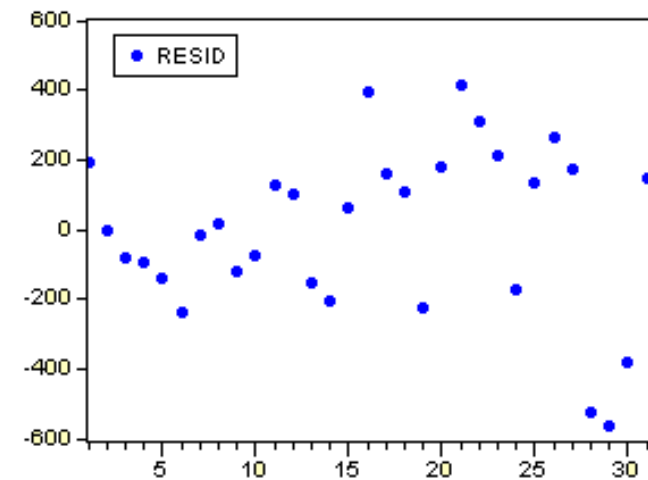
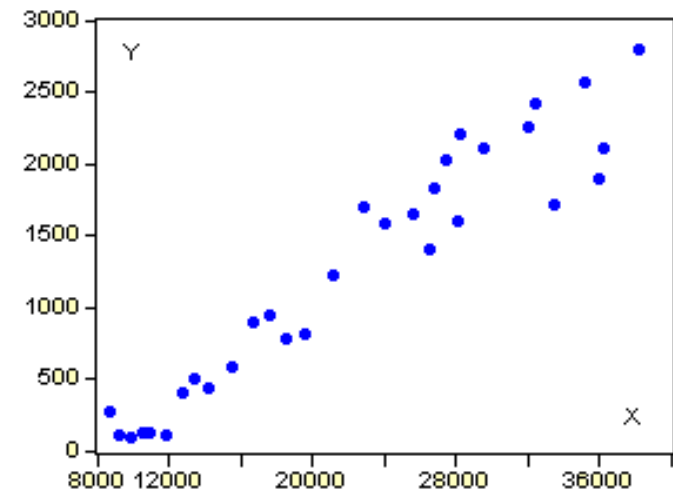
—某地区居民储蓄模型

个人储蓄（Y）与可支配（X）收入模型

Dependent Variable: Y
Method: Least Squares
Date: 02/04/07 Time: 20:41
Sample: 1 31
Included observations: 31

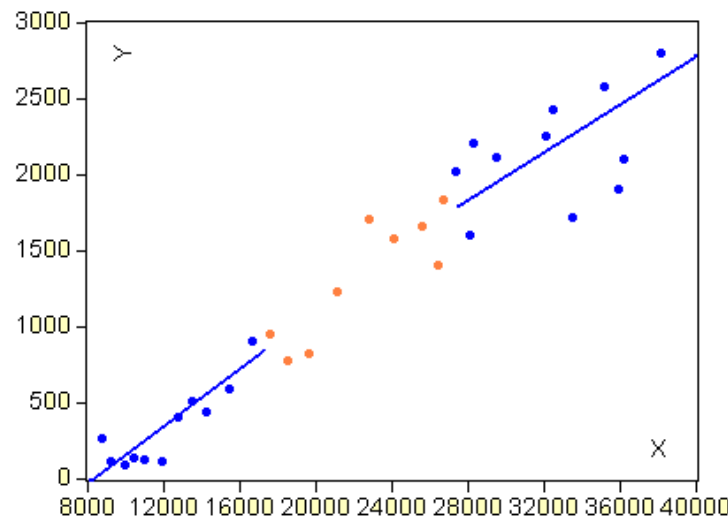
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-700.4110	116.6679	-6.003458	0.0000
X	0.087831	0.004827	18.19575	0.0000

R-squared	0.919464	Mean dependent var	1266.452
Adjusted R-squared	0.916686	S.D. dependent var	846.7570
S.E. of regression	244.4088	Akaike info criterion	13.89790
Sum squared resid	1732334.	Schwarz criterion	13.99042
Log likelihood	-213.4175	F-statistic	331.0852
Durbin-Watson stat	1.089829	Prob(F-statistic)	0.000000



个人储蓄（Y）与可支配（X）收入模型

Goldfeld-Quandt 检验



去掉中间9个观测值。

用第1个子样本回归： $\hat{Y} = -744.64 + 0.09X$ ， $SSE_1 = 150867.9$

用第2个子样本回归： $\hat{Y} = 414.98 + 0.05X$ ， $SSE_2 = 958109.4$

H_0 : u_t 具有同方差， H_1 : u_t 具有递增型异方差。

③ 构造F统计量。
$$F = \frac{SSE_2 / (n_2 - k)}{SSE_1 / (n_1 - k)} = \frac{SSE_2}{SSE_1} = \frac{958109.4}{150867.9} = 6.35$$

因为 $F = 6.35 > F_{0.05}(9, 9) = 3.18$ ，存在异方差。

个人储蓄（Y）与可支配（X）收入模型

White检验

White Heteroskedasticity Test:

F-statistic	5.819690	Prob. F(2,28)	0.007699
Obs*R-squared	9.102584	Prob. Chi-Square(2)	0.010554

更正：课本第130页丢此输出结果（图5.10）。书中图5.10应为图5.11。

White检验式

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/02/07 Time: 17:19
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19975.98	82774.93	0.241329	0.8111
X	-2.198632	8.094419	-0.271623	0.7879
X^2	0.000146	0.000176	0.830046	0.4135

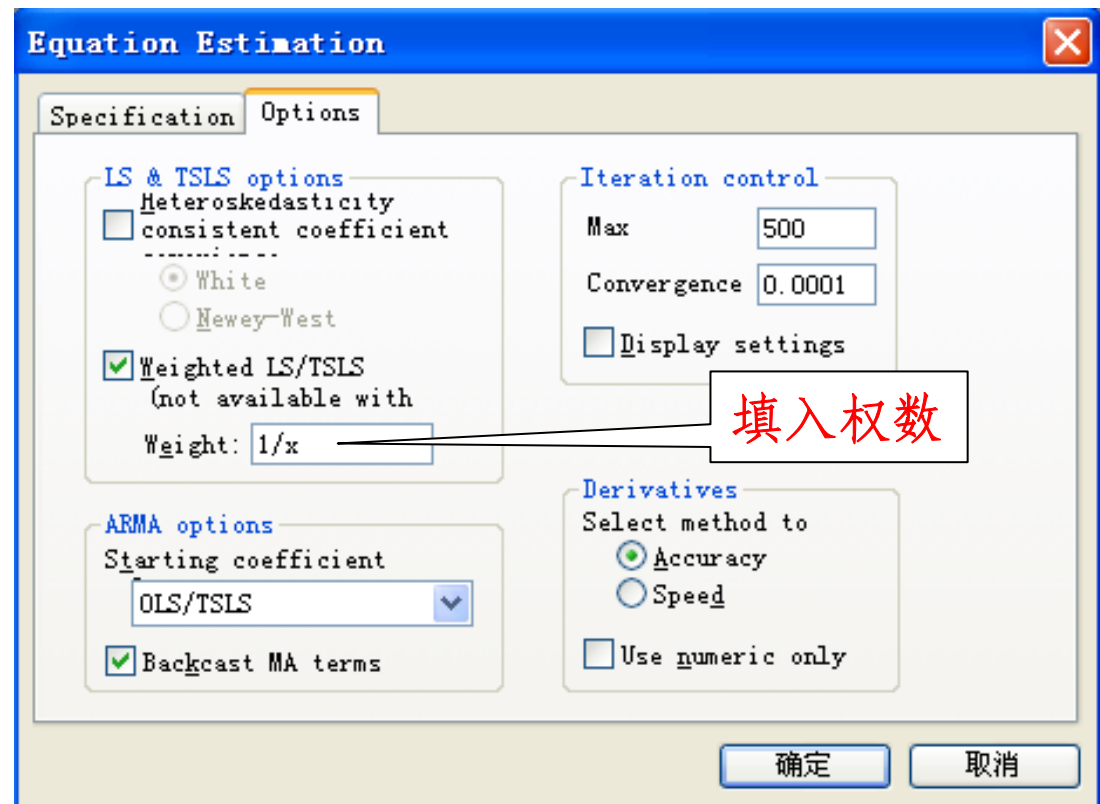
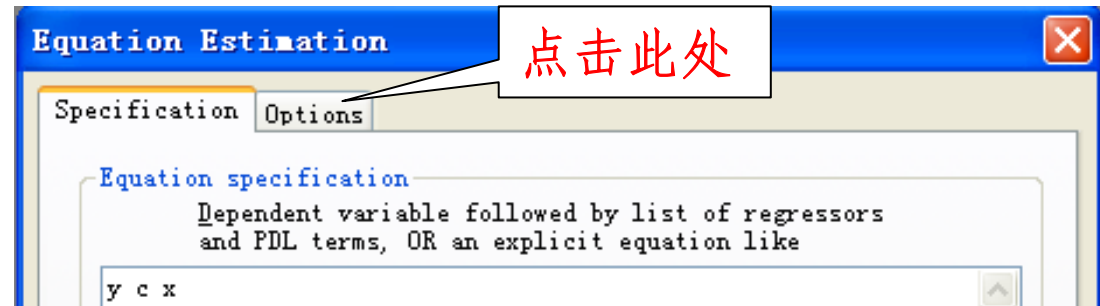
R-squared	0.293632	Mean dependent var	55881.73
Adjusted R-squared	0.243177	S.D. dependent var	77875.67
S.E. of regression	67748.39	Akaike info criterion	25.17675
Sum squared resid	1.29E+11	Schwarz criterion	25.31553
Log likelihood	-387.2397	F-statistic	5.819690
Durbin-Watson stat	1.552875	Prob(F-statistic)	0.007699

White检验的EViews操作：
在回归式窗口中点击**View**键
选**Residual Tests/ White Heteroskedasticity** 功能。
(含有无交叉项两种选择。)

(第2版教材第115页)
(第3版教材第94页)

个人储蓄（Y）与可支配（X）收入模型

加权估计（WLS）方法（1）



个人储蓄（Y）与可支配（X）收入模型

加权估计（WLS）方法（1）

Dependent Variable: Y
Method: Least Squares
Date: 02/04/07 Time: 21:54
Sample: 1 31
Included observations: 31
Weighting series: 1/X

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-742.4684	71.91567	-10.32415	0.0000
X	0.089751	0.004347	20.64696	0.0000

Weighted Statistics			
R-squared	0.786117	Mean dependent var	903.0766
Adjusted R-squared	0.778742	S.D. dependent var	406.6195
S.E. of regression	191.2661	Akaike info criterion	13.40755
Sum squared resid	1060899.	Schwarz criterion	13.50006
Log likelihood	-205.8170	F-statistic	426.2970
Durbin-Watson stat	1.081175	Prob(F-statistic)	0.000000

个人储蓄（Y）与可支配（X）收入模型

加权估计（WLS）方法（2）：用加权变量回归

Dependent Variable: Y/X
Method: Least Squares
Date: 10/02/07 Time: 18:47
Sample: 1 31
Included observations: 31

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.089751	0.004347	20.64696	0.0000
1/X	-742.4684	71.91567	-10.32415	0.0000
R-squared	0.786117	Mean dependent var	0.049255	
Adjusted R-squared	0.778742	S.D. dependent var	0.022178	
S.E. of regression	0.010432	Akaike info criterion	-6.225537	
Sum squared resid	0.003156	Schwarz criterion	-6.133022	
Log likelihood	98.49583	F-statistic	106.5881	
Durbin-Watson stat	1.081175	Prob(F-statistic)	0.000000	

White Heteroskedasticity Test:

F-statistic	3.334016	Prob. F(2,28)	0.050258
Obs*R-squared	5.962524	Prob. Chi-Square(2)	0.050729

自己把回归式还原为
Y对X回归情形。

回归系数OLS估计结果
是**0.088**，WLS估计
结果是**0.090**。**0.09**的
统计特性更好。

（不讲Spearman等级
相关系数法）

对于截面数据一定要先按解释变量排序才有可能观察到异方差

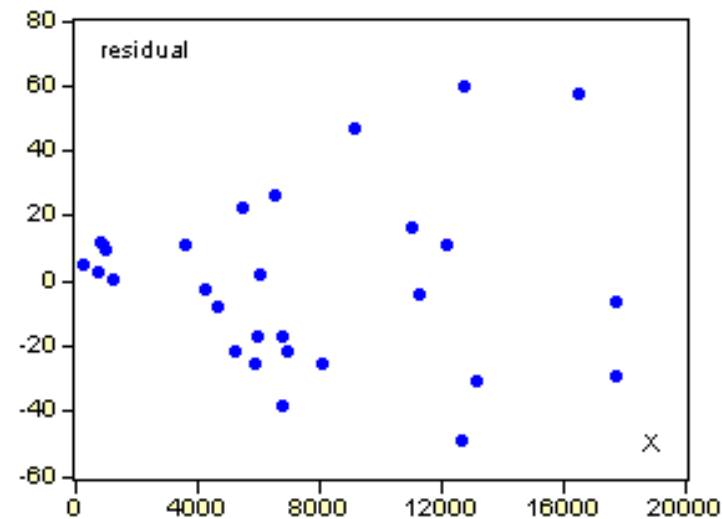
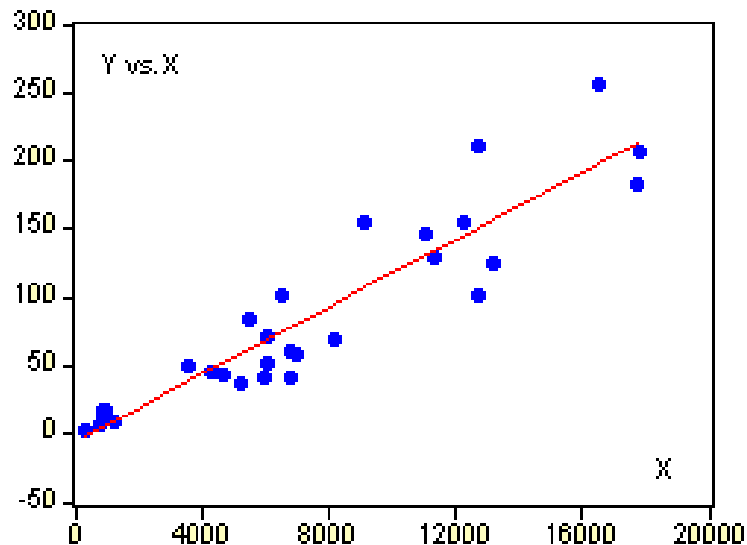
案例1：取1986年中国29个省市自治区农作物种植业产值 y_t （亿元）和农作物播种面积 x_t （万亩）数据（**file:hete01, hete02**）研究二者之间的关系。得估计的线性模型如下，

$$y_t = -5.6610 + 0.0123 x_t$$

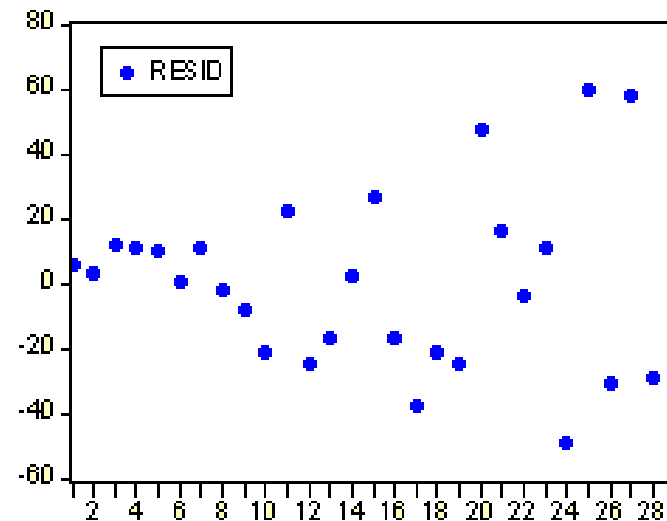
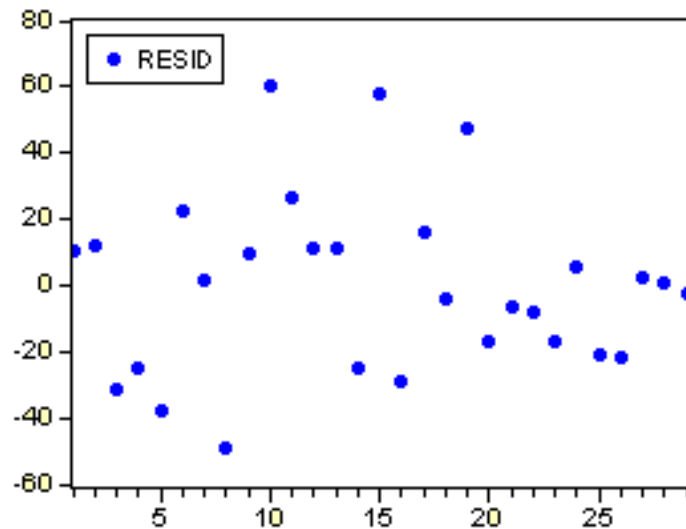
(-0.6)

(12.4)

$R^2 = 0.85, T = 29$

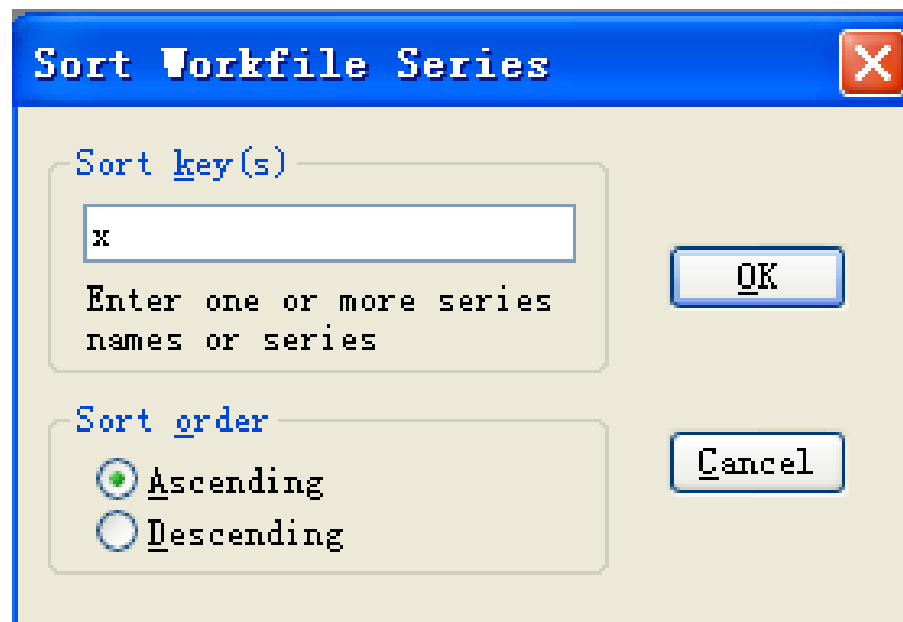


残差图中看不到异方差（左图）。原因是没有把数据按解释变量排序。数据排序并估计后得到的残差图明显存在异方差（右图）。



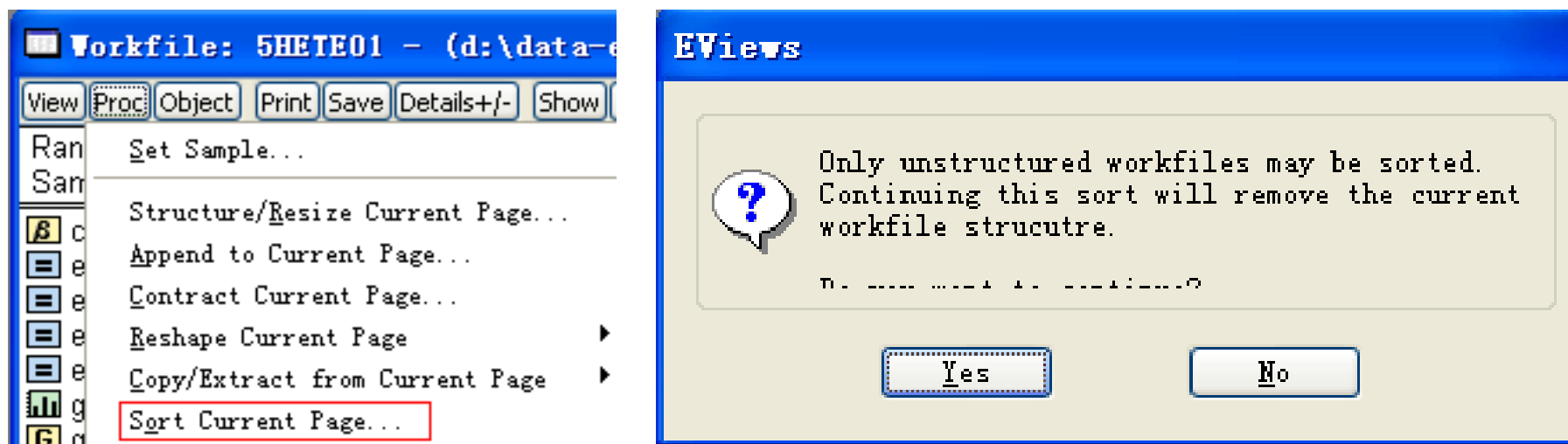
附录：用EViews 4.0给序列中的数据排序

在**Workfile**窗口点击**Procs**键并选择**Sort Series**功能，将出现一个要求填写以哪一个序列为标准（基准序列）排序的对话框。填写基准序列名，并在下侧的另一个选择框中说明是按从小到大排列（**Ascending**），还是从大到小排列（**Descending**）。缺省的选择是从小到大排列。



附录：用EViews 5.0、6.0给序列中的数据排序

在Workfile窗口点击Proc键并选Sort Curent Page功能，将出现一个警告栏。



点击Yes后，将出现一个要求填写以哪一个序列为标准（基准序列）排序的对话框。过程与EViews 4.0相同。

注意，这种操作是把工作文件中所有的变量都以选定的变量为标准排序。所以若希望保留原数据时，应先备份一个工作文件。

第一次作业（理论证明题）：

(1)：一元线性回归模型 $y = \beta_0 + \beta_1 x + u$ 中 β_1 的最小二乘估计量

用 $\hat{\beta}_1$ 表示。检验 $\beta_1 = 0$ 的 t 统计量定义为 $t = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)}$ ，其中

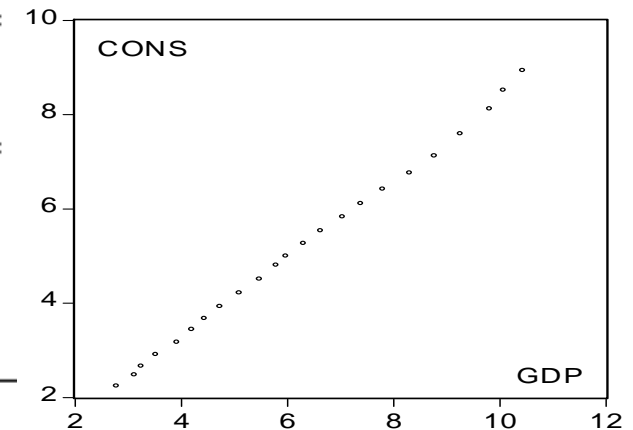
$S(\hat{\beta}_1)$ 为 $\hat{\beta}_1$ 的样本标准差。试证明统计量 $F = \frac{SSR/(2-1)}{SSE/(T-2)} = t^2$ 。

(2)：对于线性回归模型，用可决系数 R^2 表示 F 统计量。

(3) 美国1980-2002年国内总消费（**cons**：亿美元）和国内生产总值（**GDP**：亿美元）的散点图、回归结果和残差图如下：

Dependent Variable: CONS
Method: Least Squares
Date: 11/29/03 Time: 17:59
Sample: 1980 2002
Included observations: 23

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-122150.6		-2.349312	0.0287
GDP	0.843757	0.007765		0.0000
<hr/>				
R-squared		Mean dependent var	5175383.	
Adjusted R-squared	0.998140	S.D. dependent var	2009856.	
S.E. of regression		Akaike info criterion	25.66072	
Sum squared resid	1.58E+11	Schwarz criterion	25.75946	
Log likelihood	-293.0983	F-statistic		
Durbin-Watson stat	0.521514	Prob(F-statistic)	0.000000	

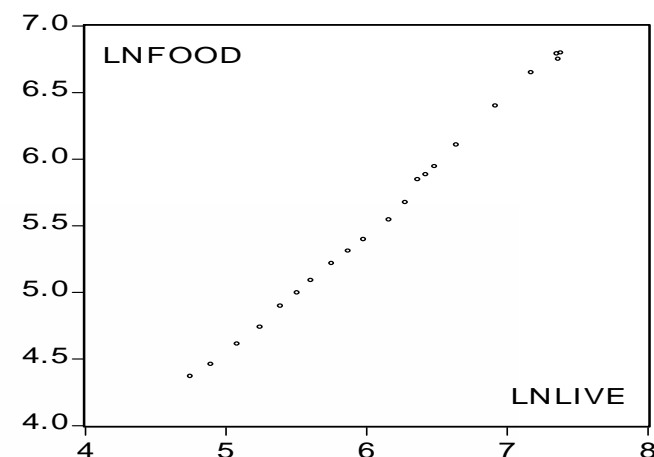


obs	Actual	Fitted	Residual	Residual Plot
1980	2232375	2236615		
1981	2468625	2519949		
1982	2651375	2627844		
1983	2899500	2860489		
1984	3159850	3196136		

- (1) 在划线处填上相应数字（共 6 处）。
- (2) 把相对于 1981 年的残差点标在残差图的相应位置上并与相邻的残差点连线。
- (3) 根据计算机输出结果，写出一元回归模型表达式。
- (4) 给定检验水平 $\alpha = 0.05$ ， F 检验的临界值 $F_{0.05}(\quad, \quad) = \underline{\hspace{2cm}}$ 。
- (5) 给定检验水平 $\alpha = 0.05$ ，单个回归参数显著性的 t 检验临界值 $t_{0.05}(\quad) = \underline{\hspace{2cm}}$ 。
- (6) 解释回归系数的经济意义。
- (7) 用 DW 统计量检验模型残差项是否存在自相关（ $\alpha = 0.05$ ）？估计残差序列的自相关系数。

(4) 中国农村居民家庭对数的人均食品支出 (*Lnfood*) 与生活支出 (*Lnlive*) 数据 (1978-1998) 散点图、一元线性回归结果如下所示。

Dependent Variable: LNFOOD
Method: Least Squares
Date: 06/22/03 Time: 10:59
Sample: 1978 1998
Included observations: 21



Variable	Coefficient	Std. Error	t-Statistic	Prob.
C		0.071924	-3.429671	0.0028
LNLIVE	0.951775	0.011631		0.0000
R-squared		Mean dependent var		5.588780
Adjusted R-squared	0.997022	S.D. dependent var		
S.E. of regression	0.042899	Akaike info criterion		-3.369525
Sum squared resid		Schwarz criterion		-3.270047
Log likelihood	37.38001	F-statistic		6696.172
Durbin-Watson stat	0.508524	Prob(F-statistic)		0.000000

- (1) 在划线处填上相应数字（共 5 处）（用保留 4 位小数运算。）
- (2) 根据以上计算机输出结果，写出估计的一元回归模型表达式。
- (3) 给定检验水平 $\alpha = 0.05$ ，检验上述回归模型显著性的临界值 $F_{0.05}(\quad, \quad) = \underline{\hspace{2cm}}$ 。
- (4) 给定检验水平 $\alpha = 0.05$ ，检验上述模型回归参数显著性的临界值 $t_{0.05}(\quad) = \underline{\hspace{2cm}}$ 。
- (5) 解释回归系数 0.951775 的经济意义。见图 1。你认为食品支出对生活支出的弹性系数值在 1978-1998 年期间有明显的变化吗？如果在 1998 以后的几年内上述弹性变小，这预示着农村居民的生活水平是提高了还是降低了？
- (6) 你认为上述回归式的误差项中存在一阶自相关吗？ 。假设存在一阶自相关，写出广义差分变量的表达式。
- (7) LM 检验辅助回归式的估计结果如下，

$$\hat{u}_t = 0.0114 - 0.0020 \ln live_t + 0.7576 \hat{u}_{t-1} - 0.1725 \hat{u}_{t-2}$$

(0.2) (-0.2) (3.1) (-0.7)

$$R^2 = 0.4251, T = 21$$

LM=? LM 检验对应的临界值是多少。试判断原回归模型误差项中是否存在二阶自相关。 ($\alpha = 0.05$)

- (8) White 检验辅助回归式的估计结果如下，

$$\hat{u}_t^2 = 0.0127 - 0.0033 \ln live_t + 0.0002 (\ln live_t)^2$$

(0.4) (-0.3) (0.3)

$$R^2 = 0.0212, T = 21$$

White 统计量=? White 检验对应的临界值是多少。试判断原回归模型误差项中是否存在异方差。 ($\alpha = 0.05$)

案例：（储蓄函数）某地区31年来居民收入与储蓄额数据表

表 4-1

单位：万元

年份	居民收入 (X)	储蓄 (Y)	年份	居民收入 (X)	储蓄 (Y)	年份	居民收入 (X)	储蓄 (Y)
1968	8777	264	1979	17663	950	1990	29560	2105
1969	9210	105	1980	18575	779	1991	28150	1600
1970	9954	90	1981	19535	819	1992	32100	2250
1971	10508	131	1982	21163	1222	1993	32500	2420
1972	10979	122	1983	22880	1072	1994	35250	2570
1973	11912	107	1984	24127	1578	1995	33500	1720
1974	12747	406	1985	25604	1654	1996	36000	1900
1975	13499	503	1986	26500	1400	1997	36200	2100
1976	14269	431	1987	27670	1829	1998	38200	2300
1977	15522	588	1988	28300	2200			
1978	16730	898	1989	27430	2017			

1、普通最小二乘估计

1、直接使用 OLS 法得:

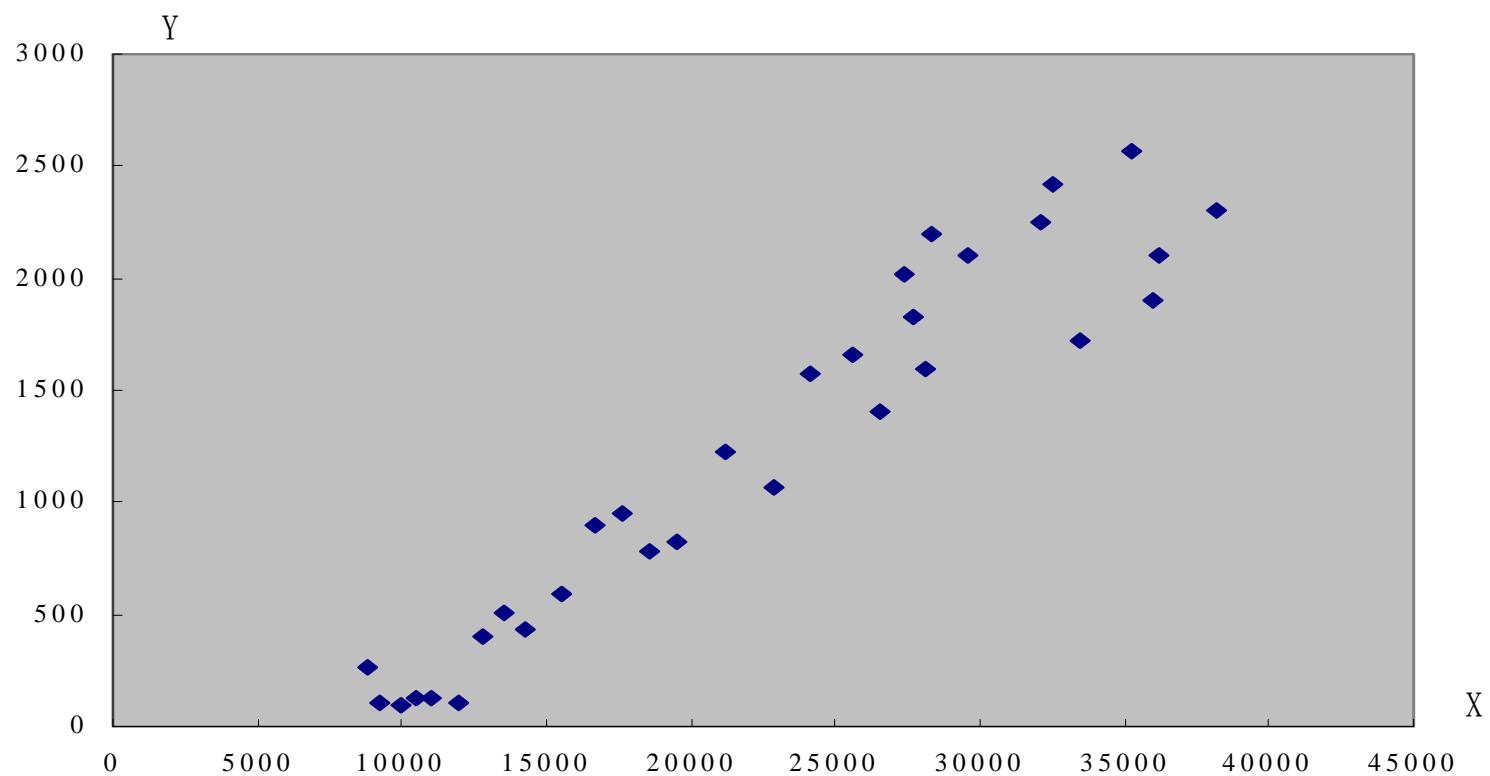
$$\hat{Y} = -665.60 + 0.0846 X$$

(-5.87) (18.04)

$$R^2=0.9182$$

2、异方差检验

(1) 图示检验



(2) G-Q检验

①求两个子样本（ $n_1=n_2=12$ ）回归方程的残差平方和RSS1与RSS2;

对第 1 个子样本（1968~1979）:

$$\hat{Y}_1 = -823.58 + 0.0954X$$

$$(-4.864) \quad (7.300)$$

$$R^2 = 0.842, \quad RSS_1 = \sum e_i^2 = 162899.2$$

对第二个子样本（1977~1988）

$$\hat{Y}_2 = 1141.07 + 0.0294X$$

$$(1.607) \quad (1.337)$$

$$R^2 = 0.1517, \quad RSS_2 = \sum e_i^2 = 769899.2$$

②计算F统计量

$$F = \text{RSS}_2 / \text{RSS}_1 = 769899.2 / 162899.2 = 4.726$$

③查表

在5%的显著性水平下，第1和第2自由度均为 $(31-7) / 2 - 2 = 10$ 的F分布临界值为

$$F_{0.05}(10, 10) = 2.97$$

由于 $F = 4.72 > F_{0.05}(10, 10) = 2.97$

因此，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**。

(3) Park检验

对直接使用 OLS 法估计的残差项的平方 \tilde{e}_i^2 进行如下一般形式的回归：

$$\ln \tilde{e}_i^2 = \alpha + \beta \ln X_i + v_i$$

得：

$$\ln \tilde{e}_i^2 = -17.99 + 2.81 \ln X_i + v_i$$

$$t \quad (-2.89) \quad (4.48)$$

$$R^2 = 0.4093$$

显然， $\ln X_i$ 前的参数表现为统计上显著的，表明原数据存在异方差性。

3、异方差模型的估计

①设异方差 $\sigma_i^2 = \sigma^2 X_i^2$,

以 $\sqrt{f(X_i)} = X_i$ 去除原模型两边, 得新模型

$$Y^* = \beta_0 X^* + \beta_1 + \mu^*$$

其中 $Y^* = Y / X_i$, $X^* = 1 / X_i$, $\mu^* = \mu / X_i$

运用 OLS 法得

$$\hat{Y}^* = -708.5 X^* + 0.086$$

$$(-10.21) \quad (20.63)$$

$$R^2 = 0.7825$$

则原模型估计为：

$$\hat{Y} = -708.5 + 0.086X$$

(-10.21) (20.63)

$$R^2 = 0.7825$$

与OLS估计结果相比较，拟合效果更差。
为什么？关于异方差形式的假定...

②如果用估计的 \tilde{e}_i^2 作为矩阵 W 的主对角线元素，即相当于用 $1/|\tilde{e}_i|$ 为权重进行加权最小二乘估计（WLS），则有

$$\hat{Y} = -686.06 + 0.0857 X$$

(-29.14) (43.59)

$$R^2 = 0.9925$$

与OLS估计结果相比较，拟合效果更好。

五、案例—2

—居民消费二元模型

1、OLS估计结果

Dependent Variable: CONS

Method: Least Squares

Date: 03/01/03 Time: 00:46

Sample: 1981 1996

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	540.5286	84.30153	6.411848	0.0000
GDP	0.480948	0.021861	22.00035	0.0000
CONS1	0.198545	0.047409	4.187969	0.0011
R-squared	0.999773	Mean dependent var	13618.94	
Adjusted R-squared	0.999739	S.D. dependent var	11360.47	
S.E. of regression	183.6831	Akaike info criterion	13.43166	
Sum squared resid	438613.2	Schwarz criterion	13.57652	
Log likelihood	-104.4533	F-statistic	28682.51	
Durbin-Watson stat	1.450101	Prob(F-statistic)	0.000000	

2、WLS估计结果

Dependent Variable: CONS

Method: Least Squares

Date: 03/01/03 Time: 00:47

Sample: 1981 1996

Included observations: 16

Weighting series: E

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	518.2881	20.52620	25.25008	0.0000
GDP	0.483814	0.003607	134.1348	0.0000
CONS1	0.193525	0.008464	22.86477	0.0000

Weighted Statistics

R-squared	0.999999	Mean dependent var	19943.81
Adjusted R-squared	0.999999	S.D. dependent var	40730.31
S.E. of regression	47.58574	Akaike info criterion	10.73030
Sum squared resid	29437.23	Schwarz criterion	10.87516
Log likelihood	-82.84243	F-statistic	980736.2
Durbin-Watson stat	1.810471	Prob(F-statistic)	0.000000

3、比较

各项统计检验指标全面改善

R^2 : 0.999739→0.999999

F: 28682→980736

Σe^2 : 438613→29437

t: 6.4 22.0 4.2→25.2 134.1 22.9

D.W.: 1.45→1.81