计量经济学II
(Econometrics II)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2022-09-25

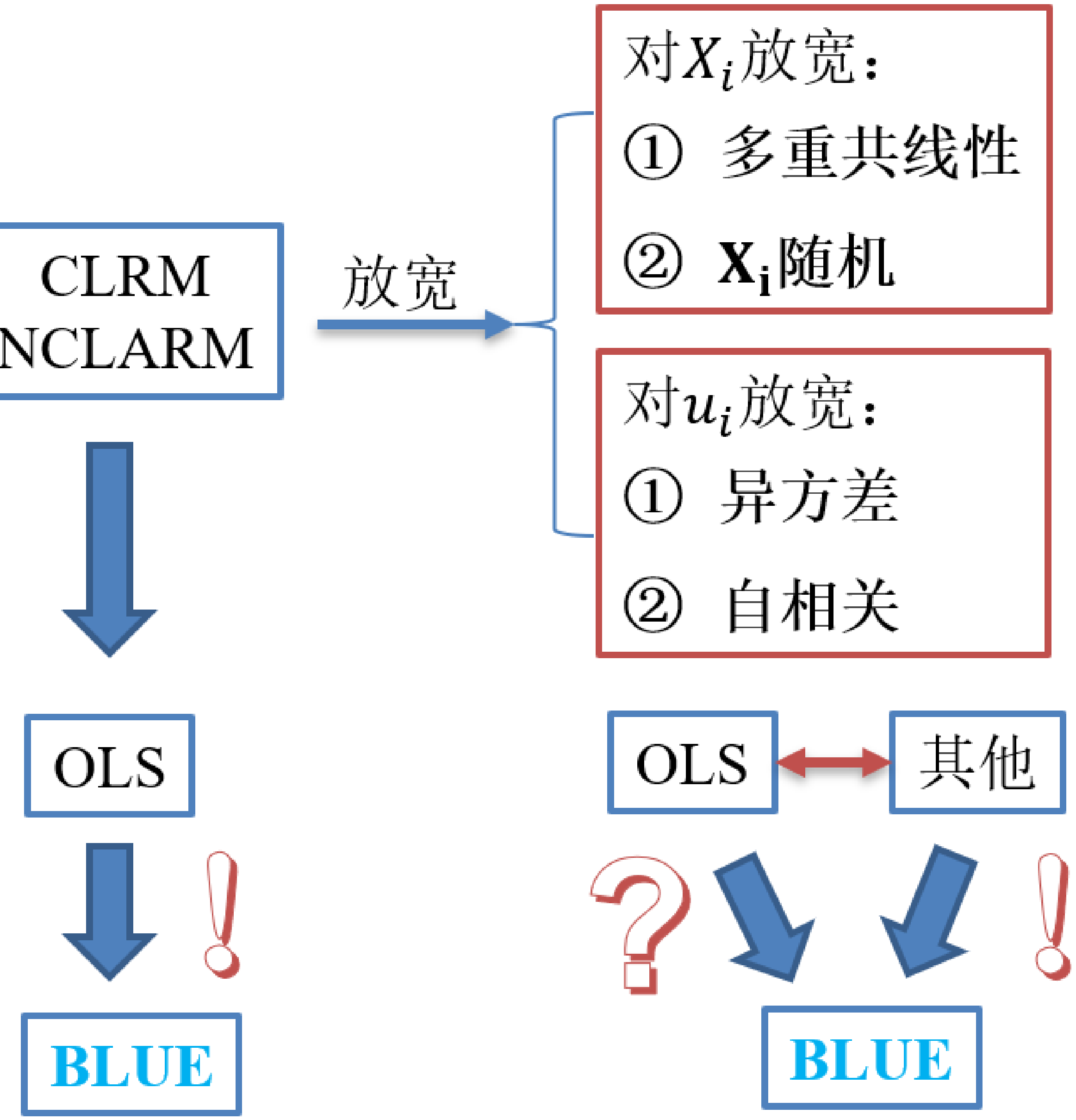
模块1：计量经济学基础

第03章 放宽假设

3.1 多重共线性(Multi-collinearity)问题

3.2 异方差(hetro-scadasticity)问题

3.3 自相关(auto-correlation)问题



定义	影响	诊断	纠正
多元回归	BLUE	辅助回归	变量变换
	LE	内生性/外生性	IV
$E(u_i^2)=\sigma_i^2$	LUE	e_i e_i^2	WLS
$E(u_t u_{t-s})\neq 0$	LUE	e_t e_{t-s}	GDE

3.2 异方差(hetro-scadasticity)问题

3.2.1 异方差性的定义和来源

k变量总体回归模型(PRM)

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i && \text{(PRM)} \\
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i && \text{(SRM)} \\
\mathbf{y} &= \mathbf{X}\beta + \mathbf{u} && \text{(PRM-matrix)} \\
\mathbf{y} &= \mathbf{X}\hat{\beta} + \mathbf{e} && \text{(SRM-matrix)}
\end{aligned}
$$

则随机干扰项的方差协方差矩阵可以写成:

$$
var - cov(\boldsymbol{u}) = E(\boldsymbol{uu}') = \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_n^2 \end{bmatrix}
$$

异方差的概念与内涵

在**CLRM假设**下，随机干扰项为**同方差**（homoscedasticity），且随机干扰项之间**不相关**。也即：

$$
\begin{aligned}
E(u_i^2) &= \sigma^2; \quad (i \in 1, 2, \cdots, n) \\
cov(u_i u_j) &= 0; \quad (i \neq j)
\end{aligned}
$$

则，随机干扰项的方差协方差矩阵进一步可以写成：

$$
var-cov(\boldsymbol{u}) = E(\boldsymbol{u}\boldsymbol{u}') \equiv \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \equiv \sigma^2 \boldsymbol{I}
$$

异方差的概念与内涵

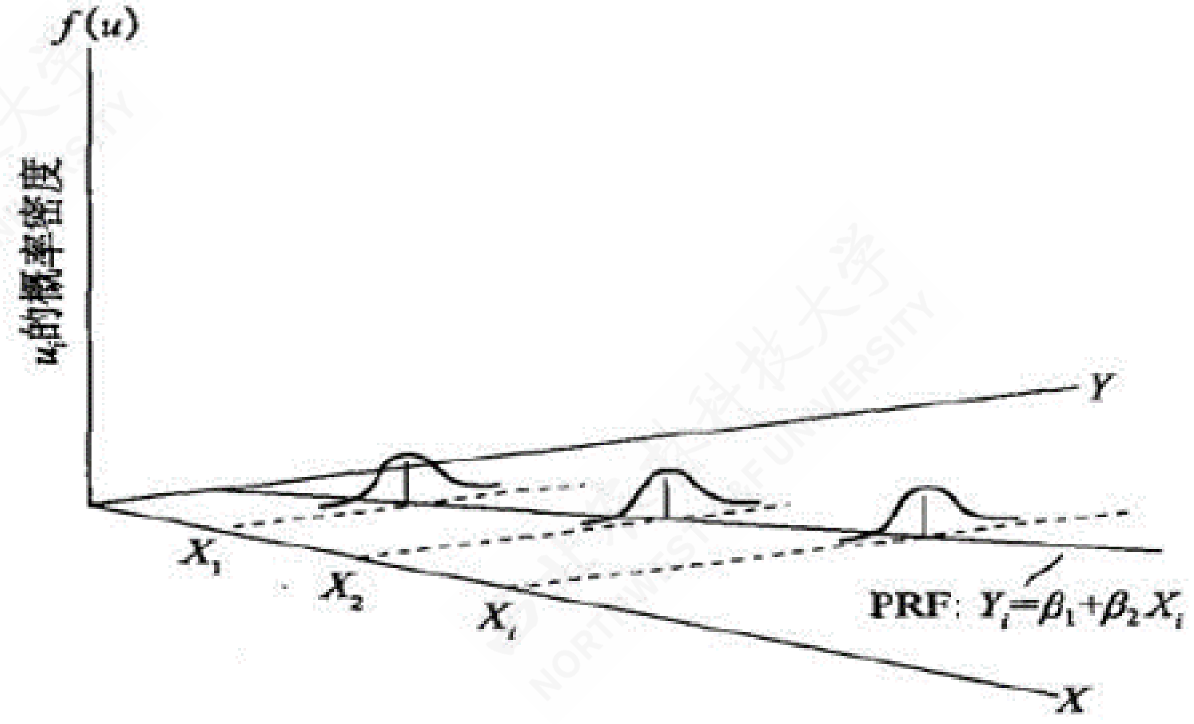
考虑违背了**CLRM假设**的一种情形：随机干扰项具有**异方差**（hetroscadasticity），但仍假设随机干扰项之间不相关。也即：

$$
\begin{aligned}
E(u_i^2) = \sigma_i^2; \quad i(\in 1,2,\cdots,n) \\
cov(u_i uj) = 0; \quad (i \neq j)
\end{aligned}
$$

随机干扰项的方差协方差矩阵可以表达为：

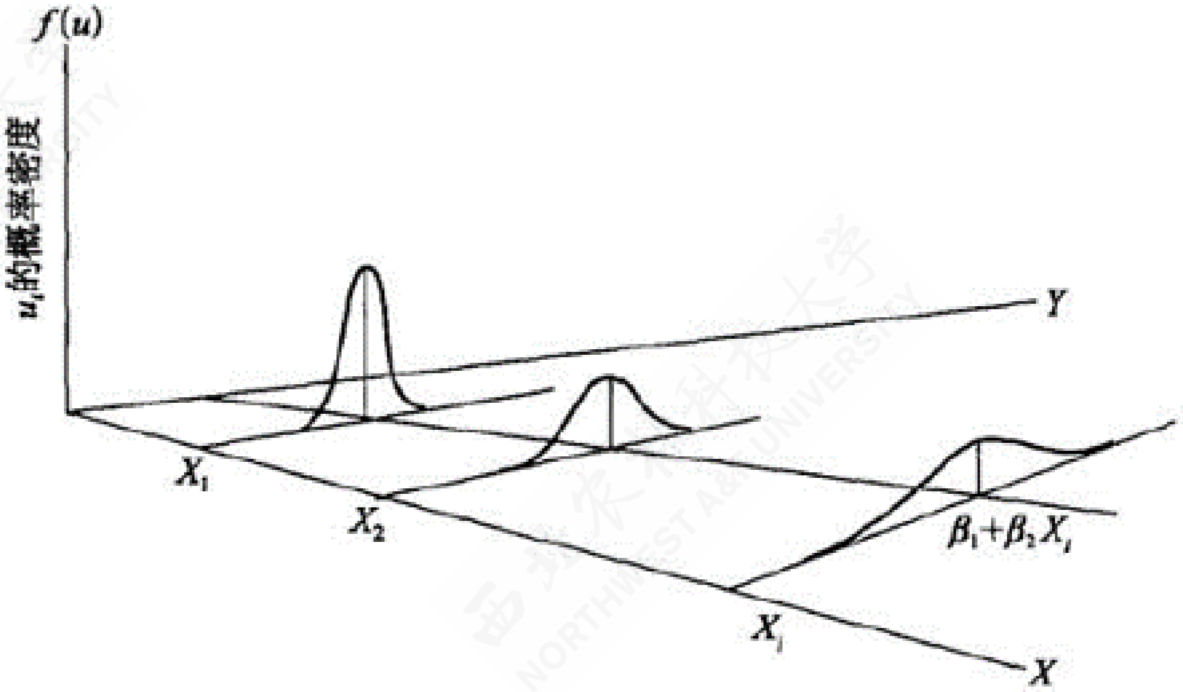
$$
var - cov(\boldsymbol{u}) = E(\boldsymbol{u}\boldsymbol{u}') = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots 0 & \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}
$$

异方差的概念与内涵(图示I)



随机干扰项的方差处处相等

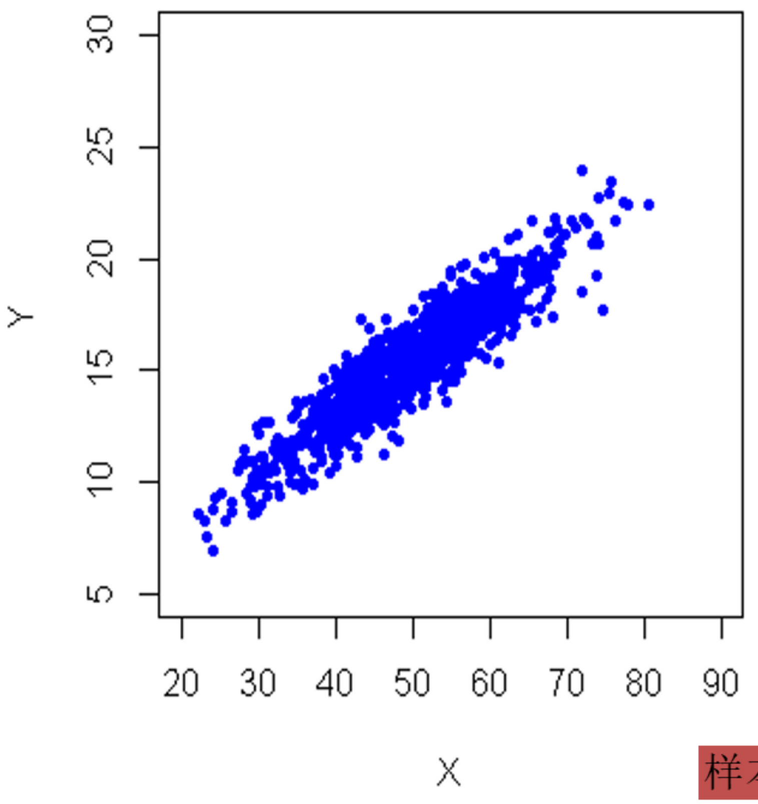
异方差的概念与内涵（图示2）



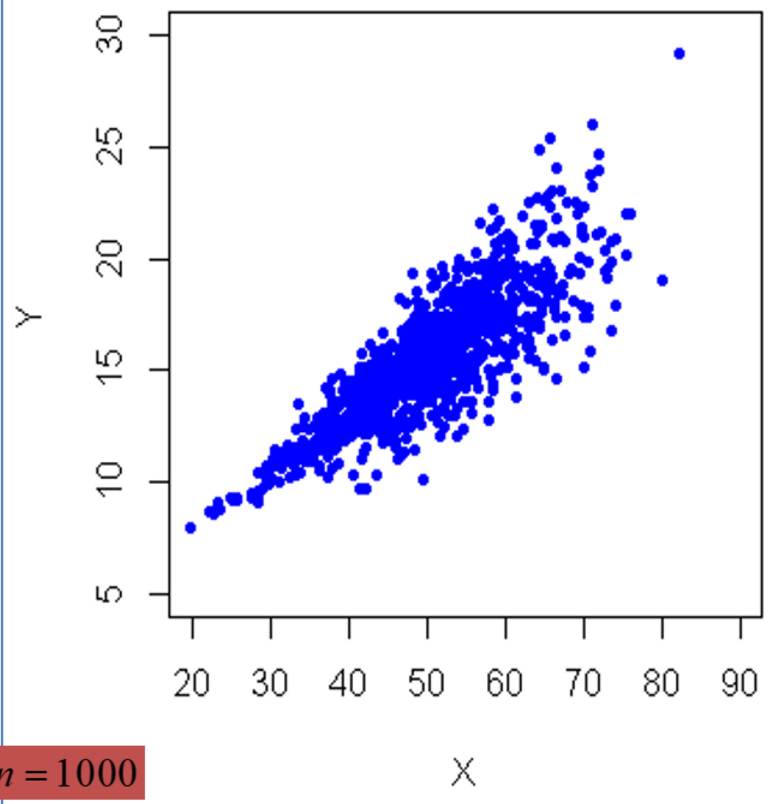
随机干扰项的方差随X取值不同而不同

数值模拟比较I

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



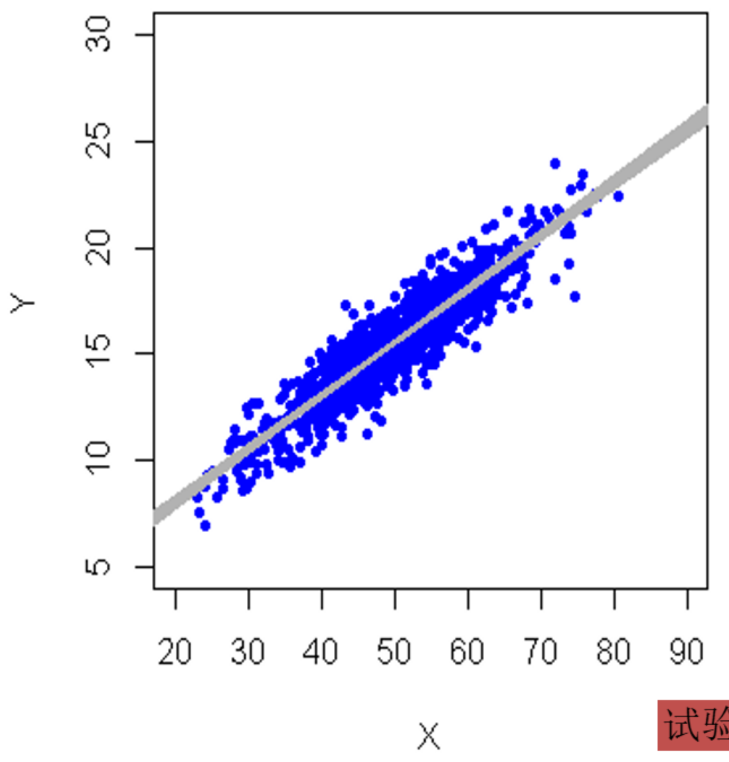
$$Y_i^* = \beta_1^* + \beta_2^* X_i + u_i^*$$



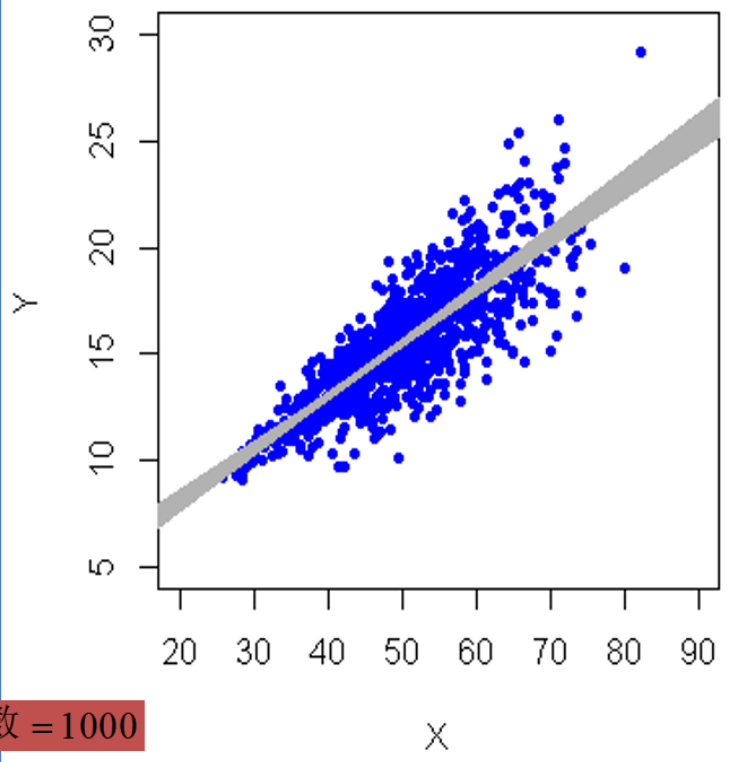
样本数$n=1000$

数值模拟比较2

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$



$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i + e_i^*$$

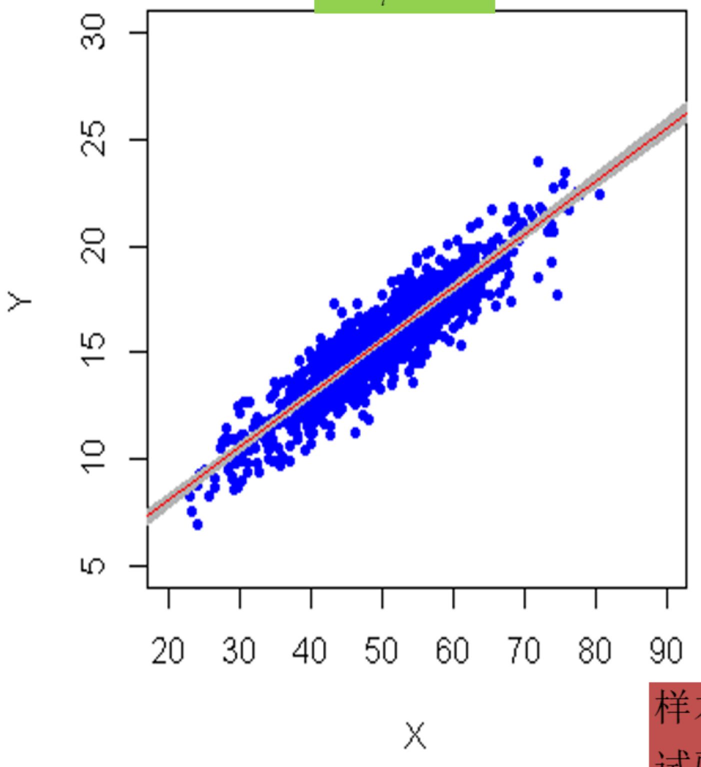


试验次数 =1000

数值模拟比较3

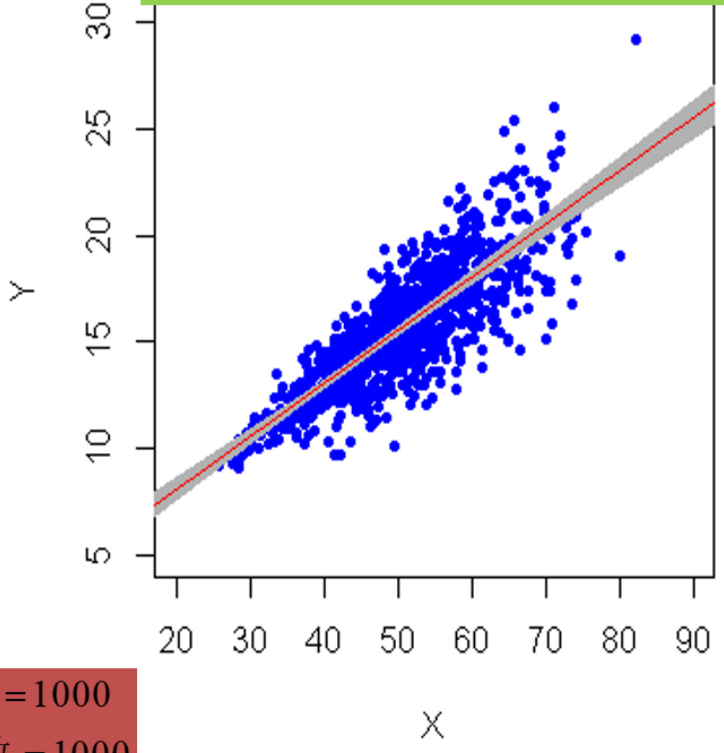
$$Y_i = 3 + 0.25X_i + u_i$$

$$\sigma_{u_i} = 1$$



$$Y_i^* = 3 + 0.25X_i + u_i^*$$

$$\sigma_{u_i^*} = 0.05 \cdot (X_i - min(X_i)) \cdot \sigma_{u_i}$$



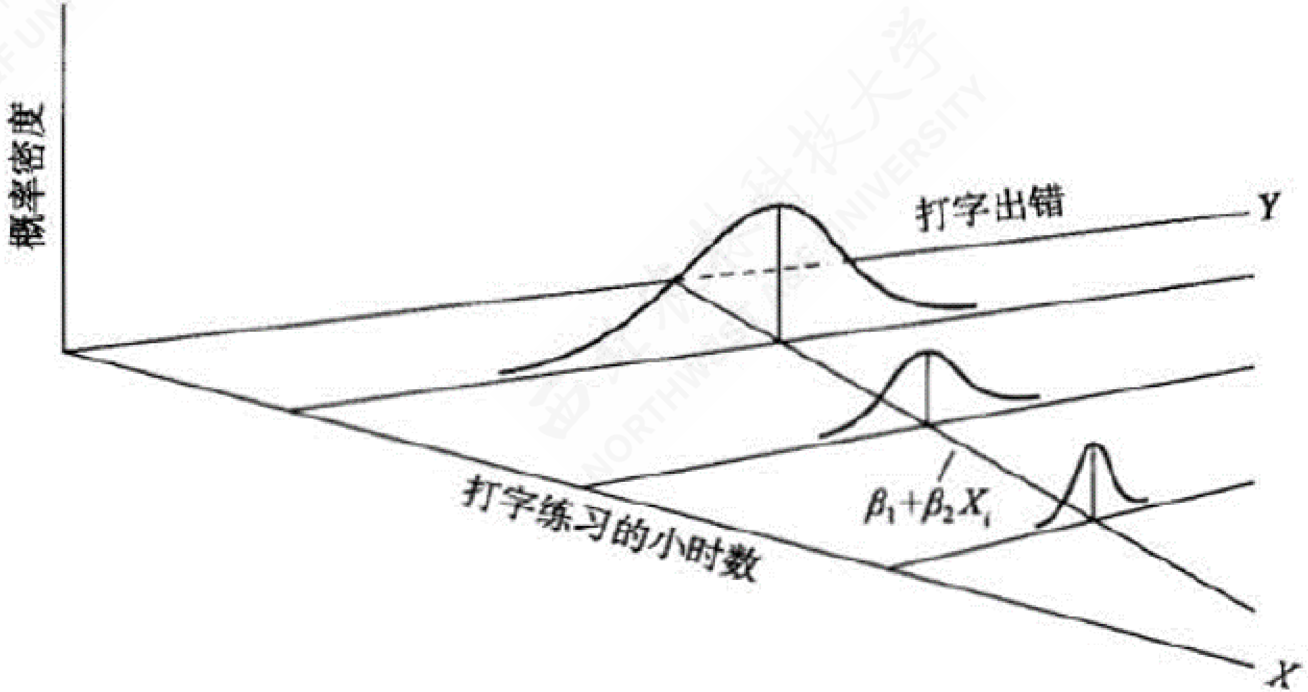
样本数 $n = 1000$

试验次数 $= 1000$

异方差的来源

异方差的来源1：边错边改误差学习模型（error-learning models）的普遍存在。

- 人们的行为误差随时间而减少。



打字出错与打字练习时长的关系

异方差的来源

异方差的来源2：随着收入的增长，人们在支出和储蓄中有更大的灵活性。

- 在做储蓄对收入的回归中，随机干扰项的方差 σ^2 与收入俱增。
- 随着收入的增长，人们在支出和储蓄中有更大的灵活性。

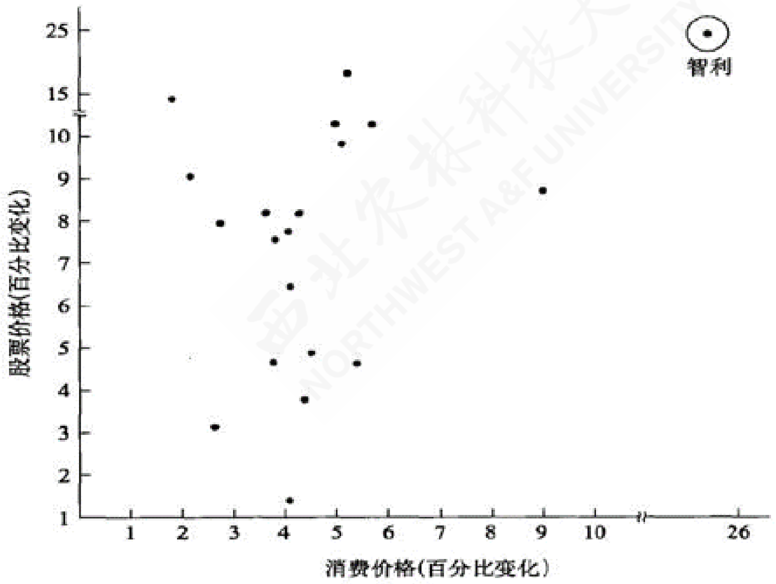
异方差的来源3：随着数据采集技术的改进，随机干扰项的方差 σ^2 可能会逐渐减小。

- 有成熟的数据处理设备的银行，在为客户提供的月度或季度报表中，相对于没有这种设备的银行，会出现更少的差错。

异方差的来源

异方差的来源4：异常值（outliers）的出现可能导致异方差性增大。

- 异常观测是来自于与产生其余观测值的总体不同的另一个总体。
- 对于小样本数据，问题会更大。

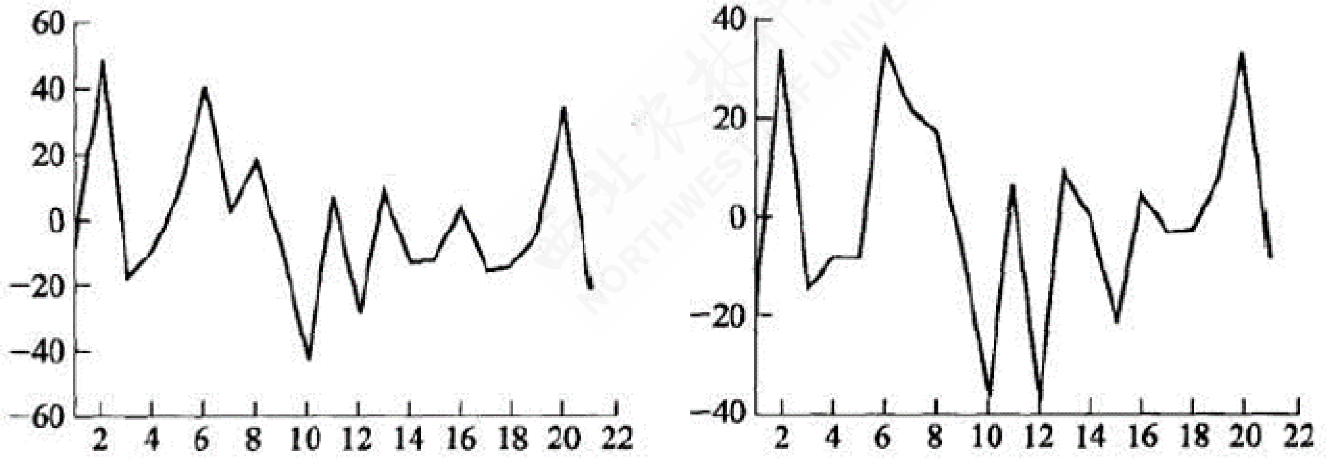


股票价格与消费价格的关系

异方差的来源

异方差的来源5：回归模型的**设定偏误**。

- 比如忽略了重要的解释变量。
- 例如，做商品的需求量对价格的回归时，没有将互补品或替代品的价格包括进来，会引起异方差问题



(a) 广告印象 Y 对广告支出 X 进行回归的残差　(b) 广告印象 Y 对 X 和 X^2 进行回归的残差

广告印象Y与广告支出X的两种回归建模结果

异方差的来源

异方差的来源6： 数据分布、数据变换和模型函数形式的原因。

- 一个或多个回归元的分布偏态(skewness)。诸如收入、财富和教育等经济变量都是很好的例子。众所周知，大多数社会中收入和财富的分配都是不匀称的，处在顶端的少数几人拥有大部分的收入和财富。
- 不正确的数据变换：(如比率或一阶差分变换等〉
- 不正确的函数形式(如线性与对数线性模型的变换)

异方差的来源

异方差的来源7：截面数据中更容易出现异方差性问题。

- 请看下面重点案例——10个行业员工薪水案例

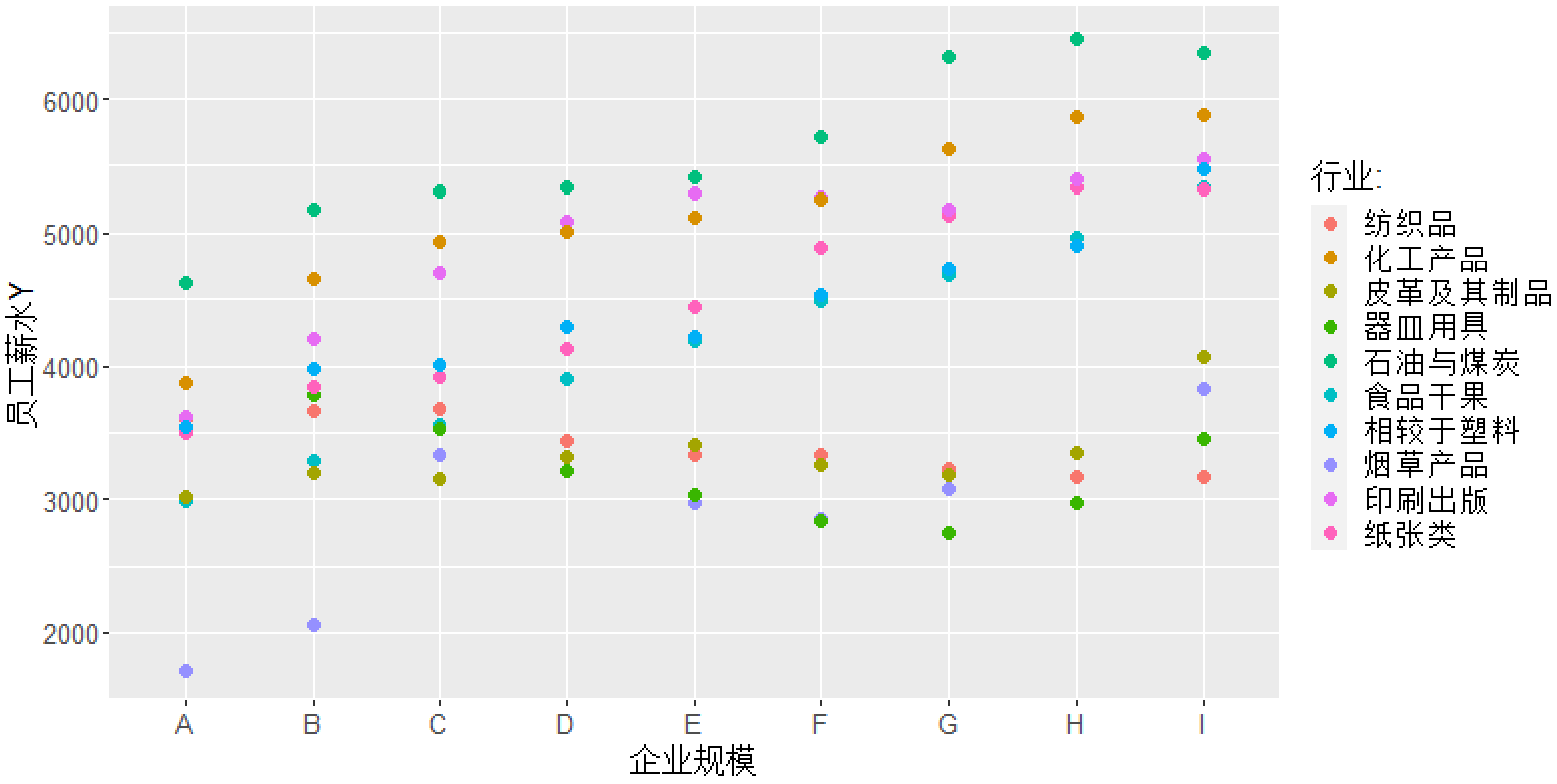
案例：行业薪水数据（原始）

Ind_chn	A	B	C	D	E	F	G	H	I
食品干果	2994	3295	3565	3907	4189	4486	4676	4968	5342
烟草产品	1721	2057	3336	3320	2980	2848	3072	2969	3822
纺织品	3600	3657	3674	3437	3340	3334	3225	3163	3168
器皿用具	3494	3787	3533	3215	3030	2834	2750	2967	3453
纸张类	3498	3847	3913	4135	4445	4885	5132	5342	5326
印刷出版	3611	4206	4695	5083	5301	5269	5182	5395	5552
化工产品	3875	4660	4930	5005	5114	5248	5630	5870	5876
石油与煤炭	4616	5181	5317	5337	5421	5710	6316	6455	6347
相较于塑料	3538	3984	4014	4287	4221	4539	4721	4905	5481
皮革及其制品	3016	3196	3149	3317	3414	3254	3177	3346	4067

10个行业、不同规模企业下员工薪水数据

其中，**1-10行**表示不同行业。**A~I列**表示不同企业规模（员工人数）：A=1-4人；B=5-9人；C=10-19人；D=20-49人；E=50-99人；F=100-249人；G=250-499人；H=500-999人；I=1000人及以上。

案例：行业薪水数据（绘图）



根据以上不同企业规模下员工薪水分布图，可以发现：

- 企业规模（人数）越大，员工薪水 Y 的分布趋向于更加分散，也即薪水分布的标准差 σ_Y 倾向于更大。

案例：行业薪水数据（汇总数据I）

根据前述原始数据，我们可以分别计算出10个行业不同企业规模下（A~I）员工的平均薪水 Y 以及不同企业规模下（A~I）员工薪水的**标准差** σ_Y。

同时，假定我们也得到了10个行业不同企业规模下员工的平均生产力 X 数据。那么，我们就可以获得如下的**汇总数据表**：

Ind_chn	A	B	C	D	E	F	G	H	I
平均薪水	3396	3787	4013	4014	4146	4241	4387	4538	4843
薪水的标准差	743.7	851.4	727.8	805.06	929.9	1080.6	1243.2	1307.7	1112.5
平均生产力	9355	8584	7962	8275	8389	9418	9795	10281	11750

10个行业、不同规模企业下员工薪水数据

其中，**A~I列**表示不同企业规模（员工人数）：A=1-4人；B=5-9人；C=10-19人；D=20-49人；E=50-99人；F=100-249人；G=250-499人；H=500-999人；I=1000人及以上。

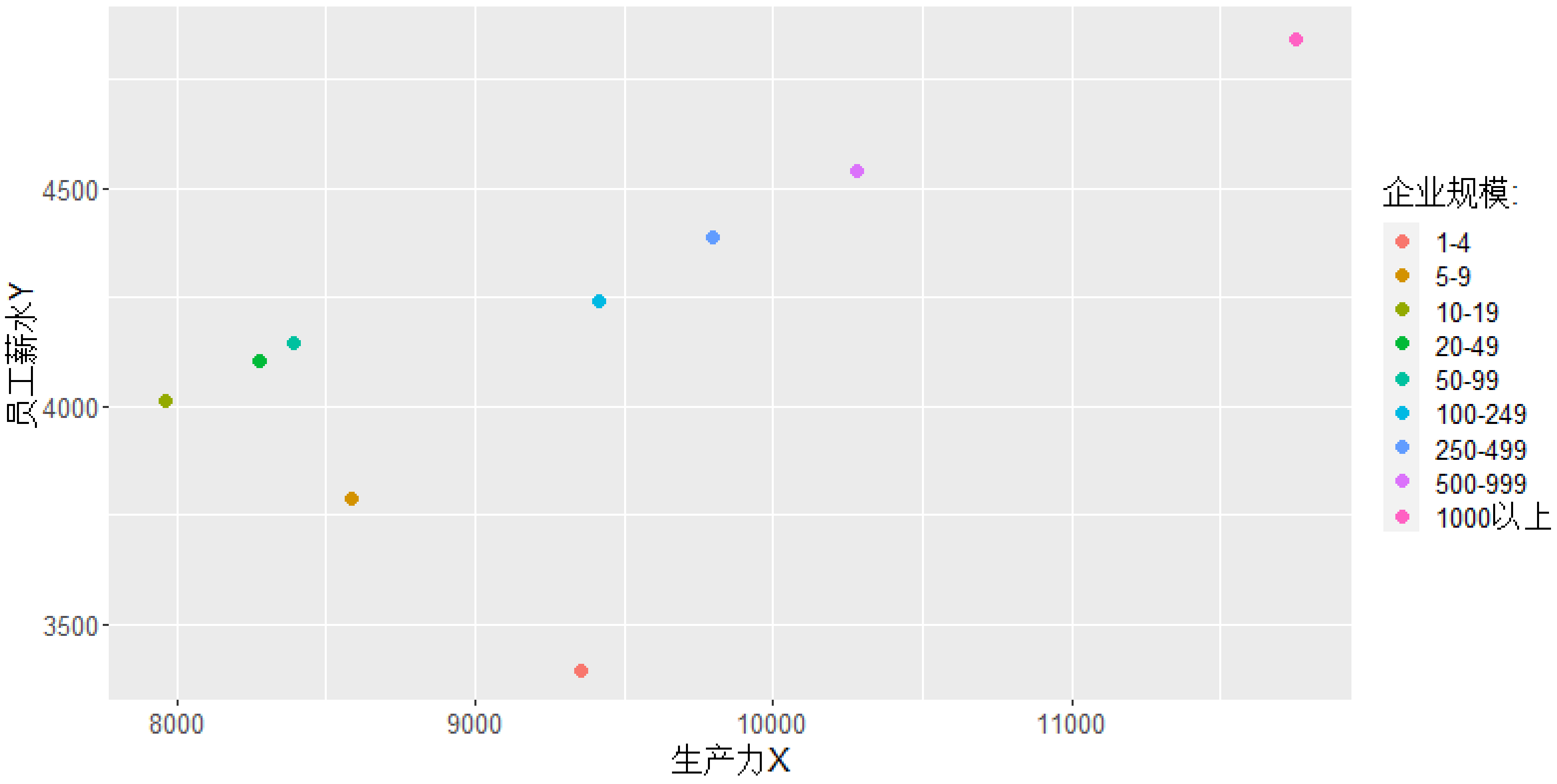
案例：行业薪水数据（汇总数据2）

我们将汇总数据表进一步整理为**标准表**：

企业规模	薪水Y	薪水标准差σ_Y	生产力X
1-4	3396.3	742.18	9355
5-9	3787	851.44	8584
10-19	4012.6	727.80	7962
20-49	4104.3	805.06	8275
50-99	4145.5	930.00	8389
100-249	4240.7	1,080.60	9418
250-499	4388.1	1,241.48	9795
500-999	4538	1,307.71	10281
1000以上	4843.4	1,110.73	11750

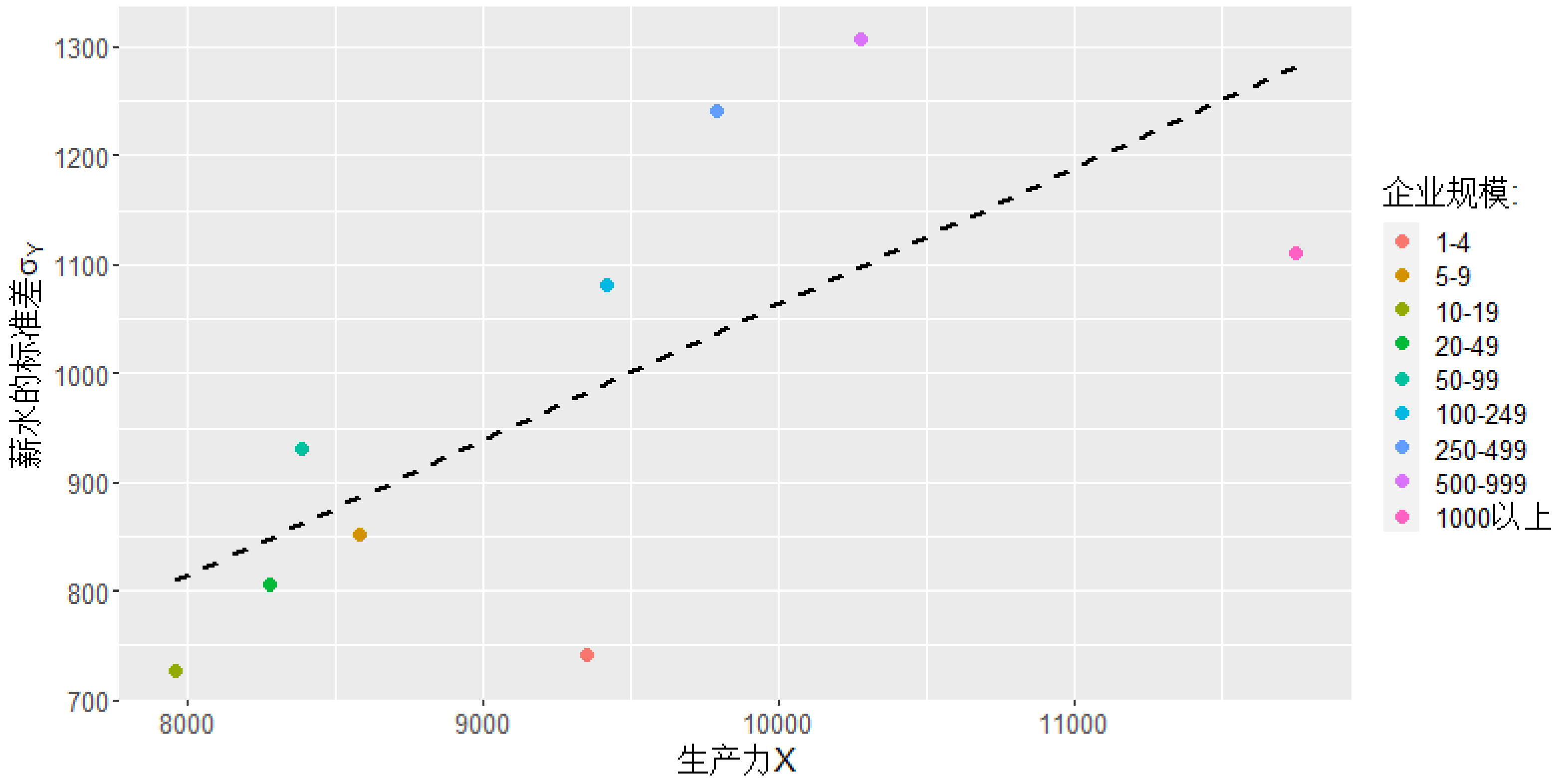
不同规模企业员工薪水与生产力数据

员工生产力 X 与员工薪水 Y 的散点图如下：



案例：行业薪水数据（汇总制图）

员工生产力 X与员工薪水标准差 σ_Y的散点图如下：



> 员工薪水标准差 σ_Y不是处处相等，而呈现明显增大（随 X增大而增大），这有违CLRM假设。

3.2.2 异方差性的影响和后果

OLS估计量不再是BLUE

对于双变量模型(一元回归模型):

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

β_2的OLS估计量 $\hat{\beta}_2$为:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

可以证明，在**CLRM假设**下OLS估计量 $\hat{\beta}_2$的方差 $\sigma^2_{\hat{\beta}_2}$ 为:

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2}$$

但是，若违背**CLRM假设**，使得随机干扰项 u_i为**异方差**，则OLS估计量 $\hat{\beta}_2$的方差 $\sigma^2_{\hat{\beta}_2}$ 为:

$$\operatorname{var}\left(\hat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

OLS估计量不再是BLUE

异方差对OLS估计带来的后果：

- 若其他CLRM假定不变，**同方差性**假定不成立时（也即随机干扰项 u_i 为**异方差**），则OLS估计量通常不再是**BLUE**。

> 也有例外，G-M定理只为OLS的有效性提供了充分条件（非必要），而OLS有效性的充要条件则由kruskal定理给出

- 此时，OLS估计量仍然是线性的和无偏的（记为**LUE**，因这两条性质都与方差无关）。但是，不再是“最优的”或“有效的”。
- 那么，在出现异方差性时，什么才是BLUE呢？

出现异方差性时的态度I

态度1：忽视异方差性，坚持错误地使用CLRM假设下OLS方法的各种公式。

在**异方差性**存在的情形下，却坚持使用OLS方法下的方差公式：

$$\hat{\beta}_2 \|_{OLS}^{\sigma^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

$$\operatorname{var}\left(\hat{\beta}_2\right) \|_{OLS}^{\sigma^2} = \frac{\sigma^2}{\sum x_i^2}$$

- 坚持使用方差公式 $\operatorname{var}\left(\hat{\beta}_2\right)\|_{OLS}^{\sigma^2}$是有偏的，可能高估或低估其真实方差 $\operatorname{var}\left(\hat{\beta}_2\right)$。
- 坚持使用回归误差方差公式 $\hat{\sigma}^2 \|_{OLS}^{\sigma^2} = \frac{\sum e_i^2}{n-2}$，并不是真值 σ^2的**无偏估计量**。
- 进一步地，置信区间、t检验和F检验也将不准确。

> 如果我们忽视异方差性而执意使用惯常的检验程序，则无论我们得出什么结论或作出什么推断，都可能产生严重的误导。

态度2：承认"**异方差性**"这一事实，但仍旧使用OLS方法。

在**异方差性**存在的情形下，直接使用OLS方法，估计量及其方差公式写成：

$$\hat{\beta}_2 \|_{OLS}^{\sigma_i^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

$$\operatorname{var}\left(\hat{\beta}_2\right) \|_{OLS}^{\sigma_i^2} = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

- 系数估计量仍是一致的
- 方差公式是有偏的，可能高估或低估其真实方差 $\operatorname{var}\left(\hat{\beta}_2\right)$。

出现异方差性时的态度3

态度3：在**异方差性**存在的情形下，首先想办法消除异方差性，再使用OLS方法。例如**加权最小二乘法**（WLS）下，估计量及其方差公式最终为写成：

$$\hat{\beta}_2^*\|_{WLS}^{\sigma_i^2} = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \qquad \leftarrow \left[w_i = \frac{1}{\sigma_i^2}\right]$$

$$\operatorname{var}\left(\hat{\beta}_2^*\right)\|_{WLS}^{\sigma_i^2} = \frac{\sum \omega_i}{\left(\sum \omega_i\right)\left(\sum \omega_i X_i^2\right) - \left(\sum \omega_i X_i\right)^2} \qquad \leftarrow \left[w_i = \frac{1}{\sigma_i^2}\right]$$

- 系数估计量是一致的
- 方差公式是无偏的，其期望将等于真实方差 $\operatorname{var}\left(\hat{\beta}_2\right)$。

出现异方差性时的态度：总结

- **态度1**："把头埋进沙堆的鸵鸟"

$$\operatorname{var}\left(\hat{\beta}_2\right) \|_{OLS}^{\sigma^2} = \frac{\sigma^2}{\sum x_i^2}$$

- **态度2**："将错就错地走下去"

$$\operatorname{var}\left(\hat{\beta}_2\right) \|_{OLS}^{\sigma_i^2} = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

- **态度3**："直面困难找出路"

$$\operatorname{var}\left(\hat{\beta}_2^*\right) \|_{WLS}^{\sigma_i^2} = \frac{\sum \omega_i}{\left(\sum \omega_i\right)\left(\sum \omega_i X_i^2\right) - \left(\sum \omega_i X_i\right)^2} \quad \leftarrow \left[w_i = \frac{1}{\sigma_i^2}\right]$$

出现异方差性时的态度：数据模拟

按照如下规则进行**蒙特卡罗模拟**设置，三种态度的差异见下表：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad \leftarrow \left[\beta_1 = 1; \beta_2 = 1; u_i \sim N\left(0, X_i^k\right); n = 20000\right]$$

k值	$\hat{\beta}_1$的标准误			$\hat{\beta}_2$的标准误		
	态度1 OLS	态度2 OLS_{het}	态度3 GLS	态度1 OLS	态度2 OLS_{het}	态度3 GLS
0.5	0.164	0.134	0.110	0.285	0.277	0.243
1.0	0.142	0.101	0.048	0.246	0.247	0.173
2.0	0.116	0.074	0.0073	0.200	0.220	0.109
3.0	0.100	0.064	0.0013	0.173	0.206	0.056
4.0	0.089	0.059	0.0003	0.154	0.195	0.017

异方差性问题的一个小结

总结：

- 异方差是指模型中关键变量 u_i 或 Y_i 行为表现“不稳定”或“不一致”。
- 异方差的表现方式多种多样，会给建模分析带来一定困扰

思考：

- 实际数据中，异方差是否很容易就能被“识别”出来？
- 识别存在异方差与判明是何种形式的异方差，那项工作更轻松？
- 薪资与生产率案例中，你能看出传统回归分析报告有什么异常么？
- 异方差问题是不是一定带来“十分显眼”的回归异常？

3.2.3 广义最小二乘法 (GLS)

广义最小二乘法 (GLS)

广义最小二乘法（Generalized Least Square，GLS）：是对普通最小二乘法（OLS）的扩展。当**CLRM假设**不能满足时，OLS方法可能无法得到**BLUE估计量**，这时常用GLS来处理其中的问题，以得到BLUE估计量。

广义最小二乘法 (GLS)

加权最小二乘法（Weighted Least Square，WLS），属于广义最小二乘法（GLS）的一种特殊情形。

加权最小二乘法（WLS）是专门用来处理随机干扰项 u_i 为异方差（违背CLRM假设，出现 $var(u_i) = \sigma_i^2$）情形的分析方法。

它通过额外的样本信息对异方差进行特定**权重**调节，从而保证随机干扰项 u_i 为**同方差**，租后获得**BLUE估计量**。

- GLS比OLS更多地利用了样本数据所提供的信息!
- GLS是更一般的方法，它包括了加权最小二乘法（WLS）、工具变量法（IV）等。
- WLS是GLS的一种特殊方法，专门用来解决异方差问题（前提是能获得**权重**信息）。
- WLS的基本思路是：对来自变异较大的总体的观测值赋予较小的权重，而对来自较小变异的总体现测值赋予较大的权重。

WLS与OLS的区别:

- OLS的思想实质是最小化:

$$\sum e_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$$

- WLS的思想实质是最小化:

$$\sum w_i e_i^2 = \sum w_i \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i\right)^2$$

其中，加权最小二乘法（WLS）需要确定**权重**向量 $\boldsymbol{\omega} = [w_1, w_2, \cdots, w_n]^t$

加权最小二乘法（WLS）：原理

对于k变量总体回归模型(PRM):

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i && \text{(PRM)} \\
Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki} + e_i && \text{(SRM)} \\
\mathbf{y} &= \mathbf{X}\beta + \mathbf{u} && \text{(PRM-matrix)} \\
\mathbf{y} &= \mathbf{X}\hat{\beta} + \mathbf{e} && \text{(SRM-matrix)}
\end{aligned}
$$

加权最小二乘法（WLS）需要确定**权重**向量 $\boldsymbol{\omega} = [w_1, w_2, \cdots, w_n]^t$，并将异方差修正为同方差 $\frac{\boldsymbol{\sigma}_i^2}{\boldsymbol{\omega}} = \boldsymbol{\sigma I}$。

$$
\begin{aligned}
var - cov(\boldsymbol{u}) &= E(\boldsymbol{u}\boldsymbol{u}') \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_2^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} = \begin{bmatrix} w_1\sigma^2 & 0 & \cdots & 0 \\ 0 & w_2\sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_n\sigma^2 \end{bmatrix}
\end{aligned}
$$

加权最小二乘法（WLS）：加权处理

对于双变量模型(一元回归模型):

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

$$\begin{aligned}
Y_i &= \beta_1 X_{0i} + \beta_2 X_i + u_i && \leftarrow [X_{0i} \equiv 1] \\
\frac{Y_i}{\sigma_i} &= \beta_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) && \leftarrow [\text{if know} \quad \sigma_i] \\
Y_i^* &= \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* && \leftarrow [\text{both divided} \quad \sigma_i]
\end{aligned}$$

$$\operatorname{var}(u_i^*) = E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 = \frac{1}{\sigma_i^2} E\left(u_i^2\right) = \frac{1}{\sigma_i^2}(\sigma_i^2) = 1$$

- 使用加权最小二乘法（WLS），转换后模型的随机干扰项满足同方差性假定，再用OLS方法，就可以得到BLUE估计量。
- 使用加权最小二乘法（WLS），得到估计量称为WLS估计量（记为 $\hat{\beta}_{WLS}$）！

加权最小二乘法（WLS）：估计量

利用样本数据，可以进行如下的WLS估计过程：

$$Y_i = \hat{\beta}_1 X_{0i} + \hat{\beta}_2 X_i + e_i \qquad \leftarrow [X_{0i} \equiv 1]$$

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \hat{\beta}_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \qquad \leftarrow [\text{if know} \quad \sigma_i]$$

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + e_i^* \qquad \leftarrow [\text{both divided} \quad \sigma_i]$$

$$\sum e_i^{*2} = \sum \left(Y_i^* - \hat{\beta}_1 X_{0i}^* - \hat{\beta}_2 X_i^*\right)^2$$

$$\sum \left(\frac{e_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1 \left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2 \left(\frac{X_i}{\sigma_i}\right)\right]^2$$

加权最小二乘法（WLS）：估计量

最小化并求偏导，得到斜率系数 $\hat{\beta}_2^*$的GLS估计量和方差：

$$\hat{\beta}_2^* = \frac{\left(\sum w_i\right)\left(\sum w_i X_i Y_i\right) - \left(\sum w_i X_i\right)\left(\sum w_i Y_i\right)}{\left(\sum w_i\right)\left(\sum w_i X_i^2\right) - \left(\sum w_i X_i\right)^2} \qquad \leftarrow \left[w_i = \frac{1}{\sigma_i^2}\right]$$

$$\operatorname{var}\left(\hat{\beta}_2^*\right) = \frac{\sum \omega_i}{\left(\sum \omega_i\right)\left(\sum \omega_i X_i^2\right) - \left(\sum \omega_i X_i\right)^2} \qquad \leftarrow \left[w_i = \frac{1}{\sigma_i^2}\right]$$

3.2.4 异方差性问题的诊断

数据类型诊断法

根据数据性质做判断（截面数据经常有）

- 在涉及不均匀、异质性（heterogeneous）单元（国家、省份、企业、家庭）的横截面数据中，异方差性可能是一种常规，而不是例外!
- 例如，投资(Y)与销售量(X2)、利率(X3)等变量之间关系的横截面分析中，如果样本同时包含小、中和大型厂家，一般都预期有异方差性。

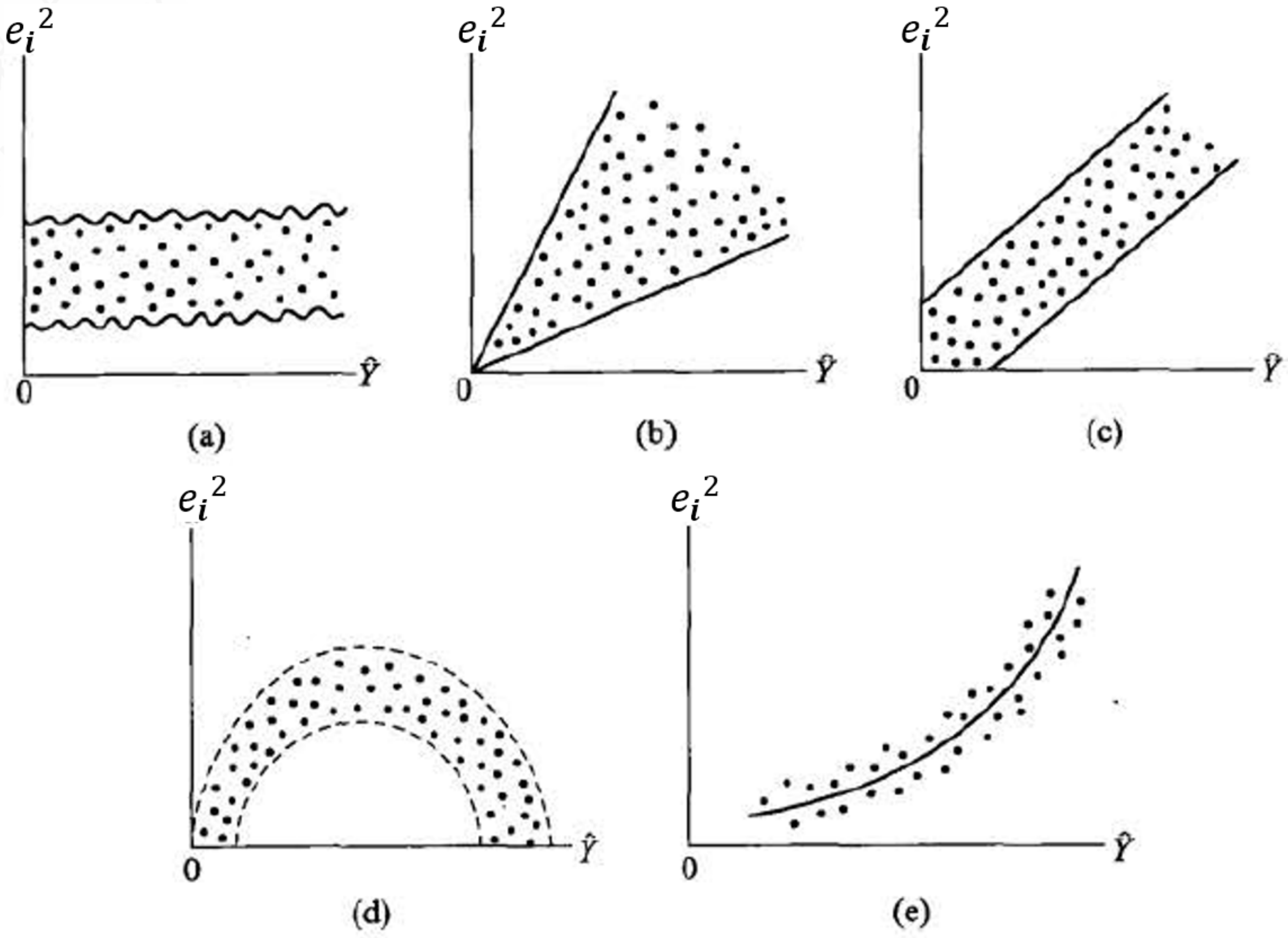
图示法

图示法重点关注模型**残差平方**序列（e_i^2）是否存在某种系统化模式。

- 图形1：残差平方序列 e_i^2 **描点图**（dot plot）。
- 图形2：残差平方序列 e_i^2 与因变量序列 Y_i 的**散点图**（scatter plot）。
- 图形3：残差平方序列 e_i^2 与自变量序列 $X_{pi},(p \in 2,3,\cdots,k)$ 的**散点图**（scatter plot）。
- 图形4：残差平方序列 e_i^2 与自变量序列平方项 $X_{pi}^2,(p \in 2,3,\cdots,k)$ 的**散点图**（scatter plot）。

图示法：模拟演示

残差平方序列 e_i^2 与因变量拟合序列 $\hat{Y}_i$ 的**散点图**的若干假想分布模式：



首先构建如下**主回归模型**：

$$Y = + \hat{\beta}_1 + \hat{\beta}_2 X + e_i$$

主回归模型回归结果为：

$$\begin{array}{llll}
\widehat{Y} = & +1990.67 & +0.23X \\
\text{(t)} & (2.1262) & (2.3358) \\
\text{(se)} & (936.2559) & (0.0998) \\
\text{(fitness)} & R^2 = 0.4380; & \bar{R}^2 = 0.3577 \\
 & F^* = 5.46; & p = 0.0522
\end{array}$$

作为对照，下面给出的是主模型的EViews报告：

Dependent Variable: Y
Method: Least Squares
Date: 04/02/19 Time: 15:32
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1990.668	936.2559	2.126200	0.0711
X	0.233148	0.099815	2.335805	0.0522

R-squared	0.438021	Mean dependent var	4161.767
Adjusted R-squared	0.357738	S.D. dependent var	420.6899
S.E. of regression	337.1460	Akaike info criterion	14.67204
Sum squared resid	795672.1	Schwarz criterion	14.71587
Log likelihood	-64.02418	Hannan-Quinn criter.	14.57746
F-statistic	5.455983	Durbin-Watson stat	0.616510
Prob(F-statistic)	0.052166		

图示法：案例演示（残差数据）

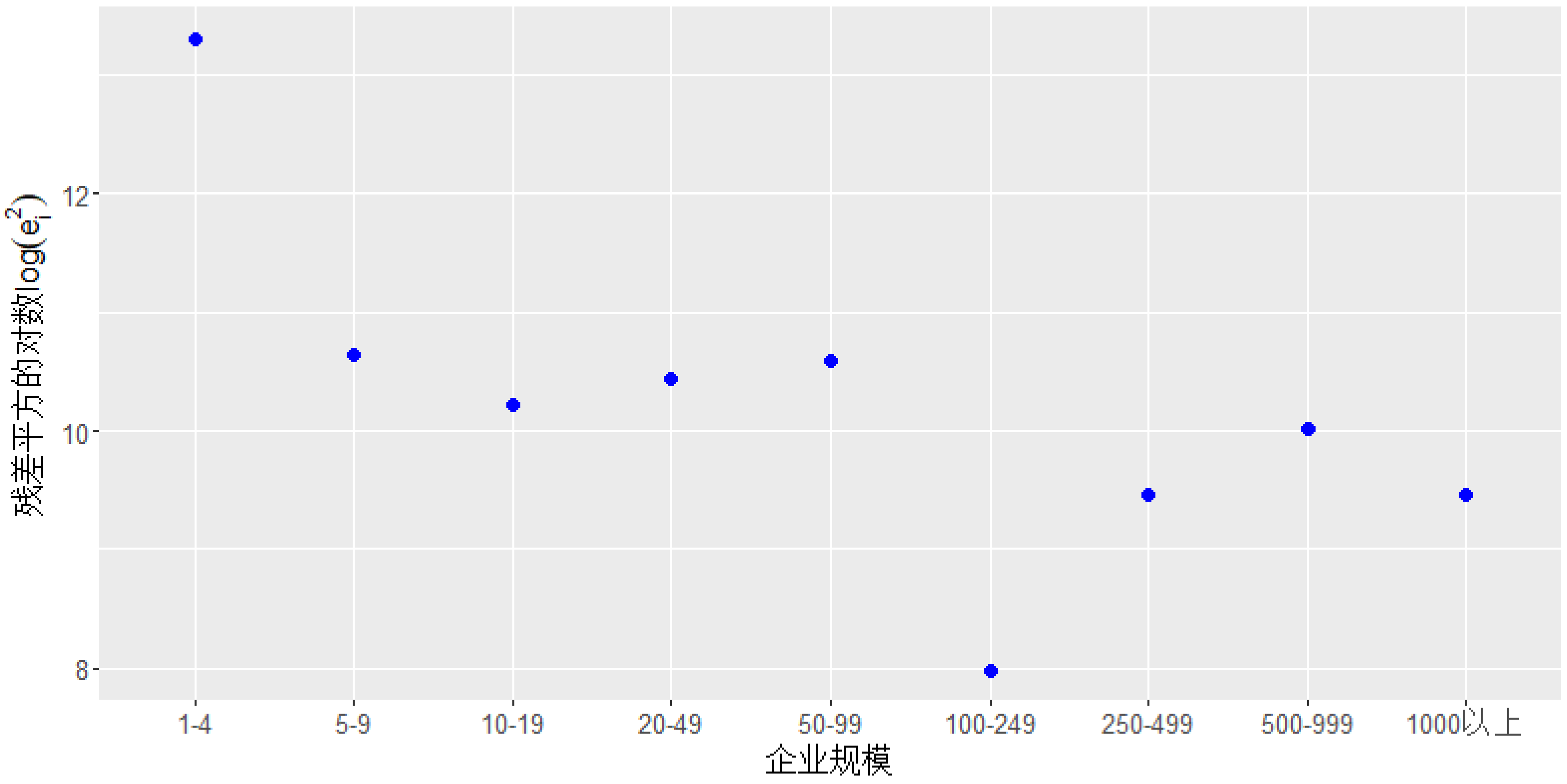
得到主回归模型的残差序列 e_i、残差平方和序列 e_i^2（及其他变换数据）：

企业规模	Y	σ_Y	X	X^2	e_i	e_i^2	$log(e_i^2)$	P_i
1-4	3396.3	742.18	9355	87516025	-775.47	601,347.70	13.31	6.80
5-9	3787	851.44	8584	73685056	-205.01	42,028.73	10.65	0.48
10-19	4012.6	727.80	7962	63393444	165.61	27,426.31	10.22	0.31
20-49	4104.3	805.06	8275	68475625	184.33	33,978.88	10.43	0.38
50-99	4145.5	930.00	8389	70375321	198.95	39,582.99	10.59	0.45
100-249	4240.7	1,080.60	9418	88698724	54.25	2,942.58	7.99	0.03

> **提示：** 后面BPJ检验中需要用到的数据 $P_i = \frac{e_i^2}{\tilde{\sigma}^2}$; $\tilde{\sigma}^2 = \frac{\sum e_i^2}{n}$

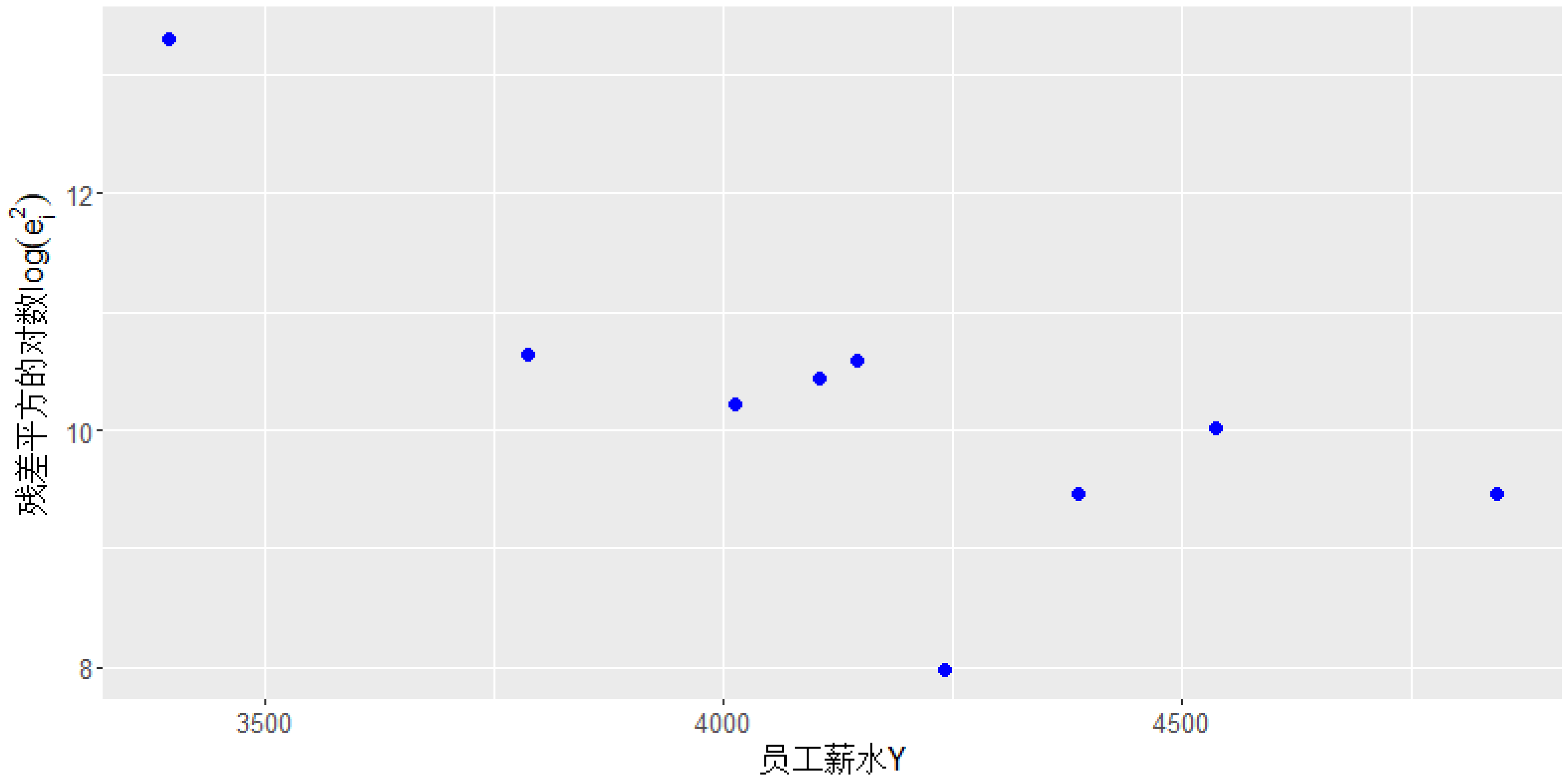
图示法：案例演示（残差分布模式I）

对数化残差平方序列 $log(e_i^2)$的**描点图**（dot plot）为：



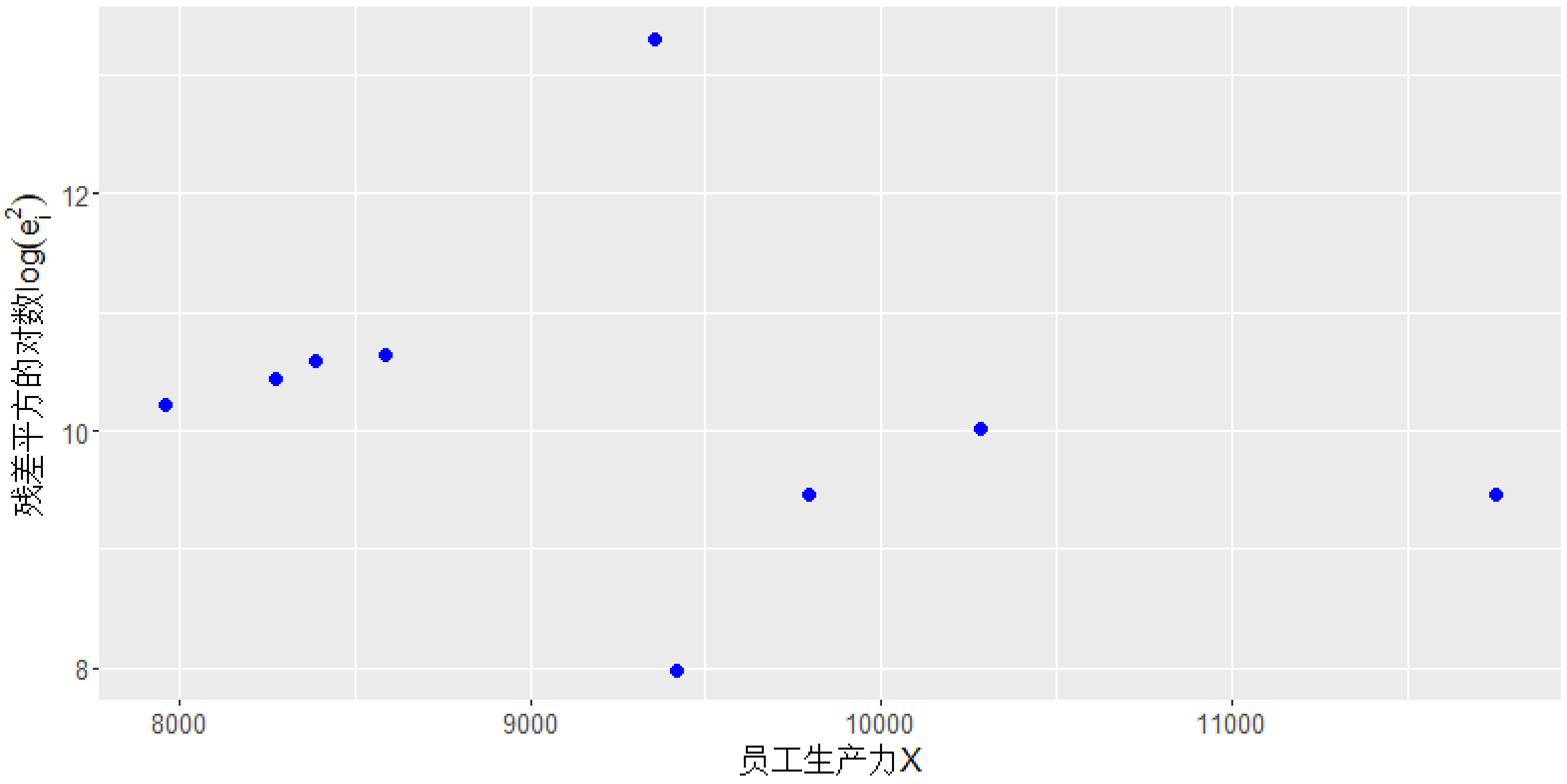
图示法：案例演示（残差分布模式2）

对数化残差平方序列 $log(e_i^2)$与 Y_i的**散点图**（scatter plot）为：



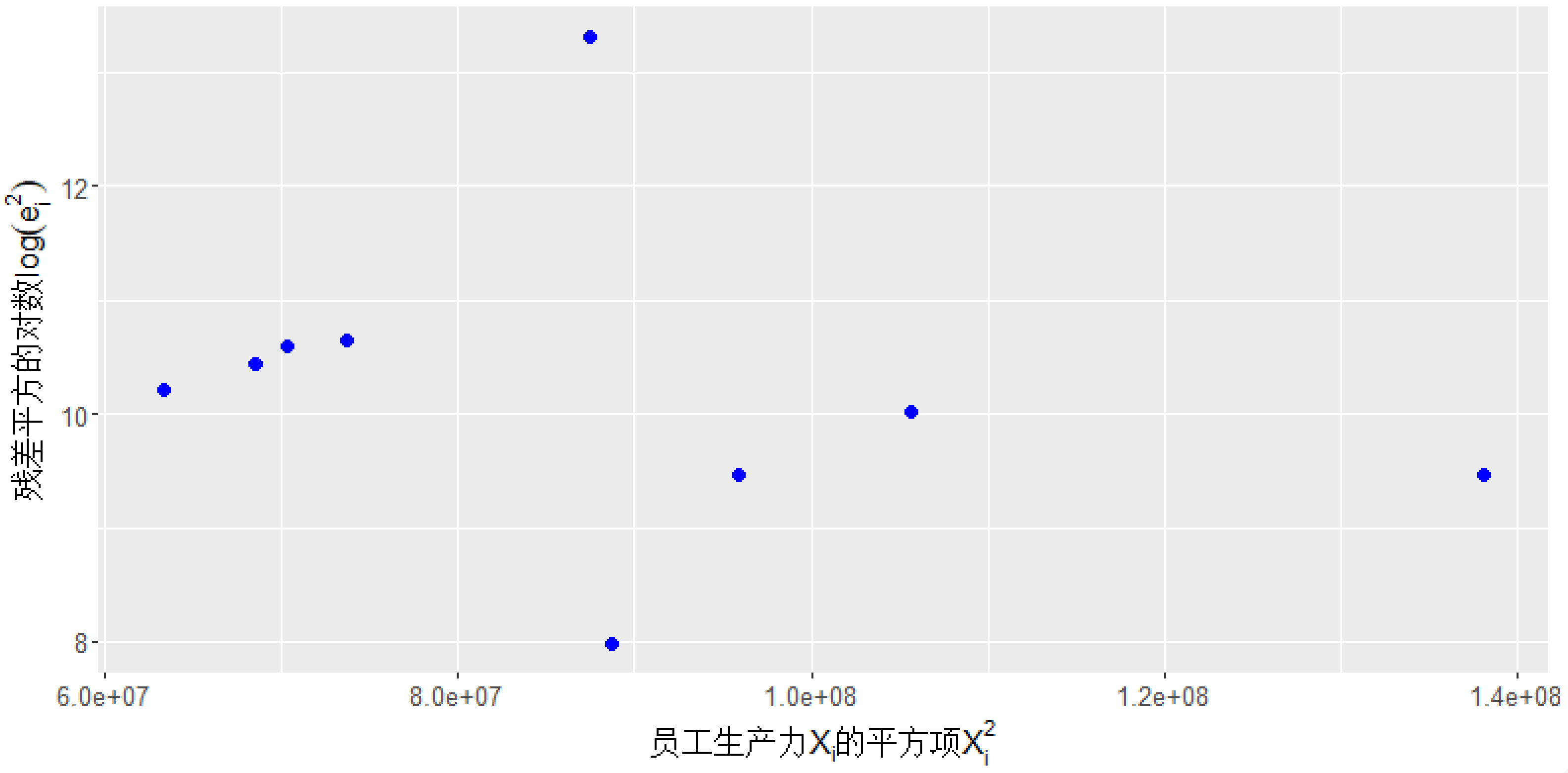
图示法：案例演示（残差分布模式3）

对数化残差平方序列 $log(e_i^2)$与 X_i的**散点图**（scatter plot）为：



图示法：案例演示（残差分布模式4）

对数化残差平方序列 $log(e_i^2)$与 X_i^2的**散点图**（scatter plot）为：



帕克检验（Park heteroscedastic test）

原理： 若 σ_i^2随着 X_i而变化（H_0 :同方差模式），则可以将其表达为：

$$
\begin{aligned}
\sigma_i^2 &= \sigma^2 X_i^{\beta} e^{v_i} \\
\ln \sigma_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i \\
\ln e_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i && \leftarrow \left[e_i^2 \simeq \sigma_i^2\right] \\
&= \alpha + \beta \ln X_i + v_i && \leftarrow \left[\alpha \equiv ln(\sigma^2)\right]
\end{aligned}
$$

帕克检验（Park heteroscedastic test）

步骤：

- 先做主回归（不考虑异方差问题，直接OLS估计）
- 再利用主回归模型的残差序列，做如下的**帕克辅助回归：**

$$ln(e_i^2) = \hat{\alpha}_1 + \hat{\alpha}_2 ln(X_{2i}) + \cdots + \hat{\alpha}_k ln(X_{ki}) + v_i$$

诊断标准：

- 如果**帕克检验辅助方程**的F检验**不显著**（对应的概率值P>0.1），则表明主模型是同方差。
- 如果**帕克检验辅助方程**的F检验**显著**（对应的概率值P<0.1），则表明主模型是异方差。

帕克检验（案例检验）

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**帕克检验辅助方程**

$$log(ei^2) = +\hat{\beta}_1 + \hat{\beta}_2 log(X) + e_i$$

帕克检验辅助方程的回归结果为：

$$\begin{array}{lll}
\widehat{log(ei^2)} = & +35.82 & -2.80 log(X) \\
(\text{t}) & (0.9334) & (-0.6667) \\
(\text{se}) & (38.3770) & (4.2021) \\
(\text{fitness}) & R^2 = 0.0597; \bar{R}^2 = -0.0746 & \\
 & F^* = 0.44; \quad p = 0.5263 &
\end{array}$$

从而可以得到**帕克检验**初步结论：F检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

薪水案例中，**帕克检验**的EViews报告如下（注意：EViews软件中没有帕克检验菜单，但可以通过哈维（Harvey）检验方法来实现）：

Heteroskedasticity Test: Harvey

F-statistic	0.444501	Prob. F(1,7)	0.5263
Obs*R-squared	0.537378	Prob. Chi-Square(1)	0.4635
Scaled explained SS	0.194271	Prob. Chi-Square(1)	0.6594

Test Equation:
Dependent Variable: LRESID2
Method: Least Squares
Date: 04/02/19 Time: 15:43
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	35.82112	38.37705	0.933399	0.3817
LOG(X)	-2.801566	4.202082	-0.666709	0.5263

R-squared	0.059709	Mean dependent var	10.23687
Adjusted R-squared	-0.074619	S.D. dependent var	1.416691
S.E. of regression	1.468596	Akaike info criterion	3.799620
Sum squared resid	15.09741	Schwarz criterion	3.843448
Log likelihood	-15.09829	Hannan-Quinn criter.	3.705040
F-statistic	0.444501	Durbin-Watson stat	1.138712
Prob(F-statistic)	0.526315		

格莱泽检验（Glejser heteroscedastic test）

原理： 若随机干扰项 $|u_i|$ 随着 X_i 而变化（ H_0 :同方差模式），则将可能的异方差模式表达为：

$$
\begin{aligned}
|u_i| &= \beta_1 + \beta_2 X_i + v_i && \text{(G1)} \\
|u_i| &= \beta_1 + \beta_2 \sqrt{X_i} + v_i && \text{(G2)} \\
|u_i| &= \beta_1 + \beta_2 \frac{1}{X_i} + v_i && \text{(G3)} \\
|u_i| &= \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i && \text{(G4)} \\
|u_i| &= \sqrt{\beta_1 + \beta_2 X_i} + v_i && \text{(G5)} \\
|u_1| &= \sqrt{\beta_1 + \beta_2 X_i^2} + v_i && \text{(G6)}
\end{aligned}
$$

格莱泽检验（Glejser heteroscedastic test）

步骤：

- 先做主回归（不考虑异方差问题，直接OLS估计）
- 再利用主回归模型的残差序列，做如下的**格莱泽辅助回归：**

$$
\begin{aligned}
|e_i| &= \beta_1 + \beta_2 X_i + v_i && \text{(G1)} \\
|e_i| &= \beta_1 + \beta_2 \sqrt{X_i} + v_i && \text{(G2)} \\
|e_i| &= \beta_1 + \beta_2 \frac{1}{X_i} + v_i && \text{(G3)}
\end{aligned}
$$

$$
\begin{aligned}
|u_i| &= \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i && \text{(G4)} \\
|e_i| &= \sqrt{\beta_1 + \beta_2 X_i} + v_i && \text{(G5)} \\
|e_1| &= \sqrt{\beta_1 + \beta_2 X_i^2} + v_i && \text{(G6)}
\end{aligned}
$$

诊断标准：

- 如**格莱泽检验辅助方程**的F检验**不显著**（对应的概率值P>0.1），则主模型是同方差。
- 如**格莱泽检验辅助方程**的F检验**显著**（对应的概率值P<0.1），则主模型是异方差。

格莱泽检验（案例检验I）

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**格莱泽检验辅助方程1**

$$abs(ei) = +\hat{\beta}_1 + \hat{\beta}_2 X + e_i$$

格莱泽检验辅助方程1的回归结果为：

$$
\begin{array}{lll}
\widehat{abs(ei)} = & +407.26 & -0.02X \\
\text{(t)} & (0.6434) & (-0.3014) \\
\text{(se)} & (632.9891) & (0.0675) \\
\text{(fitness)} & R^2 = 0.0128; \bar{R}^2 = -0.1282 & \\
 & F^* = 0.09; \quad p = 0.7719 &
\end{array}
$$

从而可以得到**格莱泽检验**初步结论：F检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

提示： `abs(ei)`表示残差的绝对值（absolute），也即 $|e_i|$。

薪水案例中，**格莱泽检验**（辅助方程1）的EViews报告如下：

Heteroskedasticity Test: Glejser

F-statistic	0.090817	Prob. F(1,7)	0.7719
Obs*R-squared	0.115270	Prob. Chi-Square(1)	0.7342
Scaled explained SS	0.114238	Prob. Chi-Square(1)	0.7354

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 04/02/19 Time: 15:36
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	407.2613	632.9891	0.643394	0.5405
X	-0.020337	0.067483	-0.301359	0.7719

R-squared	0.012808	Mean dependent var	217.8834
Adjusted R-squared	-0.128220	S.D. dependent var	214.5966
S.E. of regression	227.9396	Akaike info criterion	13.88917
Sum squared resid	363695.1	Schwarz criterion	13.93300
Log likelihood	-60.50126	Hannan-Quinn criter.	13.79459
F-statistic	0.090817	Durbin-Watson stat	1.015212
Prob(F-statistic)	0.771896		

格莱泽检验（案例检验2）

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**格莱泽检验辅助方程2**

$$abs(ei) = +\hat{\beta}_1 + \hat{\beta}_2 sqrt(X) + e_i$$

格莱泽克检验辅助方程2的回归结果为：

$$\begin{array}{lll}
\widehat{abs(ei)} = & +575.14 & -3.71 sqrt(X) \\
(\mathrm{t}) & (0.4480) & (-0.2788) \\
(\mathrm{se}) & (1283.8425) & (13.3042) \\
(\mathrm{fitness}) & R^2 = 0.0110; \bar{R}^2 = -0.1303 & \\
 & F^* = 0.08; \quad p = 0.7885 &
\end{array}$$

从而可以得到**格莱泽检验**初步结论：F检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

提示： `abs(ei)`表示残差的绝对值（absolute），也即 $|e_i|$。`sqrt(X)`表示自变量开根号（square root），也即 $\sqrt{X_i}$

格莱泽检验（案例检验2EViews报告）

薪水案例中，**格莱泽检验**（辅助方程2）的EViews报告如下：

Heteroskedasticity Test: Glejser

F-statistic	0.077708	Prob. F(1,7)	0.7885
Obs*R-squared	0.098814	Prob. Chi-Square(1)	0.7533
Scaled explained SS	0.097930	Prob. Chi-Square(1)	0.7543

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 04/02/19 Time: 15:38
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	575.1419	1283.843	0.447985	0.6677
X^0.5	-3.708701	13.30417	-0.278762	0.7885

R-squared	0.010979	Mean dependent var	217.8834
Adjusted R-squared	-0.130309	S.D. dependent var	214.5966
S.E. of regression	228.1505	Akaike info criterion	13.89102
Sum squared resid	364368.7	Schwarz criterion	13.93485
Log likelihood	-60.50958	Hannan-Quinn criter.	13.79644
F-statistic	0.077708	Durbin-Watson stat	1.011740
Prob(F-statistic)	0.788483		

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**格莱泽检验辅助方程3**

$$abs(ei) = +\hat{\beta}_1 + \hat{\beta}_2 I(1/X) + e_i$$

格莱泽克检验辅助方程3的回归结果为：

$$\begin{array}{lll}
\widehat{abs(ei)} = & +76.61 & +1297807.20 I(1/X) \\
\text{(t)} & (0.1121) & (0.2081) \\
\text{(se)} & (683.2112) & (6237225.2050) \\
\text{(fitness)} & R^2 = 0.0061; \bar{R}^2 = -0.1358 & \\
 & F^* = 0.04; \quad p = 0.8411 &
\end{array}$$

从而可以得到**格莱泽检验**初步结论：F检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

提示：`abs(ei)`表示残差的绝对值（absolute），也即 $|e_i|$。`I(1/X)`表示自变量的倒数（reciprocal），也即 $\frac{1}{X_i}$

薪水案例中，**格莱泽检验**（辅助方程3）的EViews报告如下：

Heteroskedasticity Test: Glejser

F-statistic	0.043295	Prob. F(1,7)	0.8411
Obs*R-squared	0.055323	Prob. Chi-Square(1)	0.8140
Scaled explained SS	0.054828	Prob. Chi-Square(1)	0.8149

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 04/02/19 Time: 15:39
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	76.61238	683.2112	0.112136	0.9139
1/X	1297807.	6237225.	0.208074	0.8411

R-squared	0.006147	Mean dependent var	217.8834
Adjusted R-squared	-0.135832	S.D. dependent var	214.5966
S.E. of regression	228.7072	Akaike info criterion	13.89589
Sum squared resid	366149.0	Schwarz criterion	13.93972
Log likelihood	-60.53152	Hannan-Quinn criter.	13.80131
F-statistic	0.043295	Durbin-Watson stat	0.999977
Prob(F-statistic)	0.841095		

格莱泽检验（案例检验4）

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**格莱泽检验辅助方程4**

$$abs(ei) = +\hat{\beta}_1 + \hat{\beta}_2 I(1/sqrt(X)) + e_i$$

格莱泽克检验辅助方程4的回归结果为：

$$\begin{array}{lll}
\widehat{abs(ei)} = & -91.15 & +29668.14I(1/sqrt(X)) \\
\text{(t)} & (-0.0683) & (0.2320) \\
\text{(se)} & (1334.3962) & (127897.8910) \\
\text{(fitness)} & R^2 = 0.0076; \bar{R}^2 = -0.1341 & \\
 & F^* = 0.05; \quad p = 0.8232 &
\end{array}$$

从而可以得到**格莱泽检验**初步结论：F检验不显著，不能拒绝原假设 H_0，认为**主模型**为同方差。

提示：`abs(ei)`表示残差的绝对值（absolute），也即 $|e_i|$。`I(1/sqrt(X))`表示自变量开根号的倒数，也即 $\frac{1}{\sqrt{X_i}}$

薪水案例中，**格莱泽检验**（辅助方程4）的EViews报告如下：

Heteroskedasticity Test: Glejser

F-statistic	0.053809	Prob. F(1,7)	0.8232
Obs*R-squared	0.068655	Prob. Chi-Square(1)	0.7933
Scaled explained SS	0.068041	Prob. Chi-Square(1)	0.7942

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 04/02/19 Time: 15:40
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-91.14818	1334.396	-0.068307	0.9475
X^(-0.5)	29668.14	127897.9	0.231967	0.8232

R-squared	0.007628	Mean dependent var	217.8834
Adjusted R-squared	-0.134139	S.D. dependent var	214.5966
S.E. of regression	228.5367	Akaike info criterion	13.89440
Sum squared resid	365603.2	Schwarz criterion	13.93823
Log likelihood	-60.52480	Hannan-Quinn criter.	13.79982
F-statistic	0.053809	Durbin-Watson stat	1.004082
Prob(F-statistic)	0.823199		

格莱泽检验（案例检验6）

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**格莱泽检验辅助方程6**

$$abs(ei) = +\hat{\beta}_1 + \hat{\beta}_2 I(X^2) + e_i$$

格莱泽克检验辅助方程6的回归结果为：

$$\begin{array}{llll}
\widehat{abs(ei)} = & +321.92 & -0.00 I(X^2) \\
\text{(t)} & (1.0331) & (-0.3442) \\
\text{(se)} & (311.6146) & (0.0000) \\
\text{(fitness)} & R^2 = 0.0166; \bar{R}^2 = -0.1238 & \\
 & F^* = 0.12; \quad p = 0.7408 &
\end{array}$$

从而可以得到**格莱泽检验**初步结论：F检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

提示：`abs(ei)`表示残差的绝对值（absolute），也即 $|e_i|$。`I(X^2)`表示自变量的平方，也即 X_i^2

薪水案例中，**格莱泽检验**（辅助方程6）的EViews报告如下：

Heteroskedasticity Test: Glejser

F-statistic	0.118491	Prob. F(1,7)	0.7408
Obs*R-squared	0.149809	Prob. Chi-Square(1)	0.6987
Scaled explained SS	0.148469	Prob. Chi-Square(1)	0.7000

Test Equation:
Dependent Variable: ARESID
Method: Least Squares
Date: 04/02/19 Time: 15:40
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	321.9242	311.6146	1.033085	0.3359
X^2	-1.18E-06	3.44E-06	-0.344225	0.7408

R-squared	0.016645	Mean dependent var	217.8834
Adjusted R-squared	-0.123834	S.D. dependent var	214.5966
S.E. of regression	227.4961	Akaike info criterion	13.88527
Sum squared resid	362281.2	Schwarz criterion	13.92910
Log likelihood	-60.48373	Hannan-Quinn criter.	13.79069
F-statistic	0.118491	Durbin-Watson stat	1.021291
Prob(F-statistic)	0.740787		

BPG检验（Breusch-Pagan-Goldfrey heteroscedastic test）

BPG检验有时候又被称为**BP检验**（Breusch–Pagan test），或者**BP LM检验**（Breusch–Pagan Lagrange Multiplier test）。

原理：若 σ_i^2随着某些**非随机变量** Z_{mi}而变化（ H_0 :同方差模式），则可以将其表达为：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \\
\sigma_i^2 &= f\left(\alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \cdots + \alpha_m Z_{mi}\right) \\
&= \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \cdots + \alpha_m Z_{mi}
\end{aligned}
$$

其中，部分或全部的自变量 $X_{pi}, p \in (2, 3, \cdots, k)$可以作为 Z_{mi}。

BPG检验（Breusch-Pagan-Goldfrey heteroscedastic test）

步骤：

- 先做主回归（不考虑异方差问题，直接OLS估计）
- 再利用主回归模型的残差序列，做如下的**BPG辅助回归：**

$$P_i = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \cdots + \alpha_m Z_{mi} \qquad \leftarrow \left[P_i = \frac{e_i^2}{\tilde{\sigma}^2}; \quad \tilde{\sigma}^2 = \frac{\sum e_i^2}{n} \right]$$

- 计算**BPG辅助回归**方程的ESS(回归平方和)，并得到如下**LM统计量**（卡方统计量）：

$$LM \equiv \chi^{2^*} = \frac{ESS}{2} \sim \chi^2(m-1)$$

BPG检验（Breusch-Pagan-Goldfrey heteroscedastic test）

诊断标准：

- 如果**BPG辅助回归方程**的卡方检验**不显著**，也即 $LM \equiv {\chi^2}^* < \chi^2_{1-\alpha}(m-1)$（对应的概率值P>0.1），则表明主模型是同方差。
- 如果**BPG辅助回归方程**的卡方检验**显著**，也即 $LM \equiv {\chi^2}^* > \chi^2_{1-\alpha}(m-1)$（对应的概率值P<0.1），则表明主模型是异方差。

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**BPG检验辅助方程**

$$Pi = +\hat{\beta}_1 + \hat{\beta}_2 X + e_i$$

BPG检验辅助方程的回归结果为：

$$
\begin{array}{llll}
\widehat{Pi} = & +1.50 & -0.00X \\
\text{(t)} & (0.2314) & (-0.0774) \\
\text{(se)} & (6.4714) & (0.0007) \\
\text{(fitness)} & R^2 = 0.0009; & \bar{R}^2 = -0.1419 \\
 & F^* = 0.01; & p = 0.9404
\end{array}
$$

BPG检验辅助方程的回归平方和为 ESS = -5.3979；**LM统计量**（绝对值）为 $LM \equiv \chi^{2^*} = ESS/2 = 2.6990$。给定显著性水平 $\alpha = 0.05$时卡方分布的理论查表值为 $\chi^2_{1-\alpha}(m-1) = \chi^2_{0.95}(1) = 3.841459$。从而可以得到**BPG检验**初步结论：卡方检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

> **提示**： $P_i = \frac{e_i^2}{\tilde{\sigma}^2}$； $\tilde{\sigma}^2 = \frac{\sum e_i^2}{n}$

BPG检验（案例检验EViews报告）

薪水案例中，**BPG检验**的EViews报告如下（**注意**：EViews采用了BPG不同分析方法，EViews结果（本页）与手动计算结果（上一页））：

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.005998	Prob. F(1,7)	0.9404
Obs*R-squared	0.007706	Prob. Chi-Square(1)	0.9300
Scaled explained SS	0.009853	Prob. Chi-Square(1)	0.9209

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/02/19 Time: 15:44
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	132398.3	572120.7	0.231417	0.8236
X	-4.723990	60.99413	-0.077450	0.9404

R-squared	0.000856	Mean dependent var	88408.01
Adjusted R-squared	-0.141879	S.D. dependent var	192797.4
S.E. of regression	206020.8	Akaike info criterion	27.50247
Sum squared resid	2.97E+11	Schwarz criterion	27.54630
Log likelihood	-121.7611	Hannan-Quinn criter.	27.40789
F-statistic	0.005998	Durbin-Watson stat	1.072440
Prob(F-statistic)	0.940433		

GQ检验（ Goldfeld–Quandt heteroscedastic test ）

原理： 若 σ_i^2随着某一个自变量的平方项 $(X_i^2$而变化（ H_0 :同方差模式），则可以将其表达为：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_2 + u_i \\
\sigma_i^2 &= \sigma^2 X_i^2
\end{aligned}
$$

GQ检验（ Goldfeld–Quandt heteroscedastic test ）

步骤：

- 把数据按 X_i 由小到大排序，从小到大；去掉中间大约c个数（预先确定的）
- 把数据分成两份样本：前一半样本数 $(n-c)/2$，后一半样本数 $(n-c)/2$
- 对两份**分段样本**数据，做如下的**GQ辅助回归**，并分别进行OLS回归

$$
\begin{aligned}
\overleftarrow{Y_i} &= \overleftarrow{\alpha_1} + \overleftarrow{\alpha_2}\overleftarrow{X}_i + \overleftarrow{e_i} \\
\overrightarrow{Y_i} &= \overrightarrow{\alpha_1} + \overrightarrow{\alpha_2}\overrightarrow{X}_i + \overrightarrow{e_i}
\end{aligned}
$$

- 分别得到两个辅助回归方程的残差平方和：RSS_1和 RSS_2，其分别对应的自由度为 $df_1 = df_2 = (n-c-2k)/2$。
- 计算得到如下F统计量：

$$
F^* = \frac{\mathrm{RSS}_1/df_1}{\mathrm{RSS}_2/df_2} \sim F((n-c-2k)/2, (n-c-2k)/2)
$$

GQ检验（Goldfeld–Quandt heteroscedastic test）

诊断标准：

- 如果**GQ辅助回归方程**的F检验**不显著**，也即 $F^* < F_{1-\alpha}((n-c-2k)/2,(n-c-2k)/2)$（对应的概率值P>0.1），则表明主模型是同方差。
- 如果**GQ辅助回归方程**的F检验**显著**，也即 $F^* > F_{1-\alpha}((n-c-2k)/2,(n-c-2k)/2)$（对应的概率值P<0.1），则表明主模型是异方差。

怀特检验（White heteroscedastic test）

步骤：

- 先做主回归（不考虑异方差问题，直接OLS估计）
- 再利用主回归模型的残差序列，做如下的**怀特辅助回归**：

$$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$$

- 计算**怀特辅助回归**方程的判定系数 R^2，并得到如下卡方统计量（$\tilde{k}$为辅助方程中回归系数个数）：

$$\chi^* = n \cdot R^2 \sim \chi^2(\tilde{k} - 1)$$

怀特检验（White heteroscedastic test）

诊断标准：

- 如果**怀特辅助回归方程**的卡方检验**不显著**，也即 $\chi^* < \chi^2_{1-\alpha}(m-1)$（对应的概率值 P>0.1），则表明主模型是同方差。
- 如果**怀特辅助回归方程**的卡方检验**显著**，也即 $\chi^* < \chi^2_{1-\alpha}(m-1)$（对应的概率值 P<0.1），则表明主模型是异方差。

薪水案例中，为验证**主模型**是否存在异方差问题。我们可以构建如下的**怀特检验辅助方程**

$$I(ei^2) = +\hat{\beta}_1 + \hat{\beta}_2 X + \hat{\beta}_3 I(X^2) + e_i$$

怀特检验辅助方程的回归结果为：

$$
\begin{array}{llll}
\widehat{I(ei^2)} = & -3793851.07 & +810.57X & -0.04I(X^2) \\
\text{(t)} & (-0.7835) & (0.8105) & (-0.8169) \\
\text{(se)} & (4842167.3561) & (1000.0435) & (0.0510) \\
\text{(fitness)} & R^2 = 0.1008; & \bar{R}^2 = -0.1989 & \\
 & F^* = 0.34; & p = 0.7269 &
\end{array}
$$

怀特检验辅助方程的判定系数为 $R^2=0.1008$；样本卡方统计量（绝对值）为 $\chi^{2^*}=nR^2=0.9076$。给定显著性水平 $\alpha=0.05$时卡方分布的理论查表值为 $\chi^2_{1-\alpha}(m-1)=\chi^2_{0.95}(1)=5.991465$。从而可以得到**怀特检验**初步结论：卡方检验不显著，不能拒绝原假设 H_0，认为**主模型**为**同方差**。

> **提示：** `I(ei^2)`表示残差的平方，也即 e_i^2。`I(X^2)`表示自变量平方，也即 X_i^2

怀特检验（案例检验EViews报告）

薪水案例中，**怀特检验**的EViews报告如下：

Heteroskedasticity Test: White

F-statistic	0.336482	Prob. F(2,6)	0.7269
Obs*R-squared	0.907644	Prob. Chi-Square(2)	0.6352
Scaled explained SS	1.160547	Prob. Chi-Square(2)	0.5597

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/02/19 Time: 15:41
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3793851.	4842167.	-0.783503	0.4631
X^2	-0.041665	0.051007	-0.816854	0.4452
X	810.5686	1000.043	0.810533	0.4486

R-squared	0.100849	Mean dependent var	88408.01
Adjusted R-squared	-0.198868	S.D. dependent var	192797.4
S.E. of regression	211099.3	Akaike info criterion	27.61925
Sum squared resid	2.67E+11	Schwarz criterion	27.68499
Log likelihood	-121.2866	Hannan-Quinn criter.	27.47738
F-statistic	0.336482	Durbin-Watson stat	1.120832
Prob(F-statistic)	0.726938		

3.2.5 异方差性问题的矫正

矫正情形1： 方差 σ_i^2 已知且等于样本方差 $S_{Y_i}^2$

如果主模型存在异方差问题，且假设方差正比于样本方差 $S_{Y_i}^2$，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 S_{Y_i}^2$$

对主回归方程两边同时除以 S_i，得到：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \\ \frac{Y_i}{S_i} &= \frac{\beta_1}{S_i} + \beta_2 \frac{X_{2i}}{S_i} + \cdots + \beta_k \frac{X_{ki}}{S_i} + \frac{u_i}{S_i} \\ Y_i^* &= \beta_1^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + v_i \end{aligned}$$

容易证明新模型将消除异方差性：

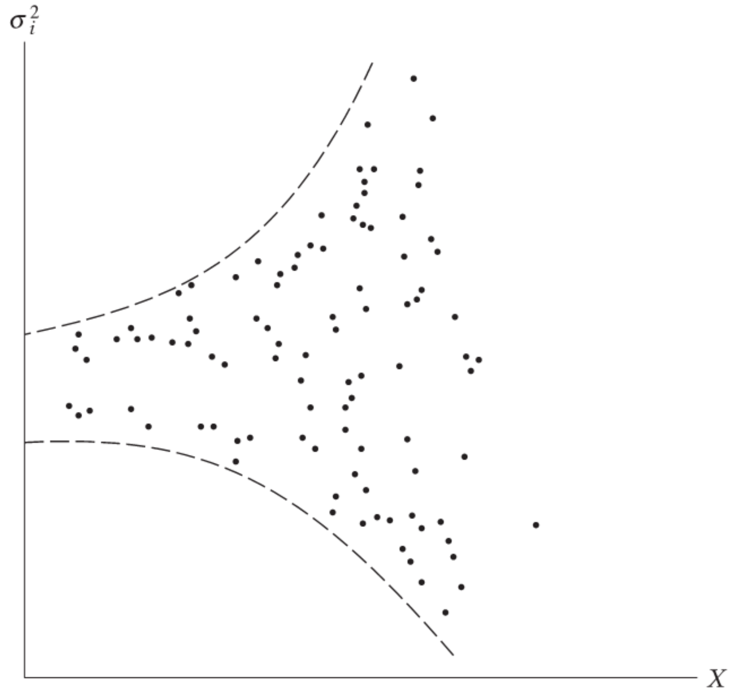
$$E\left(v_i^2\right) = E\left(\frac{u_i}{S_{Y_i}}\right)^2 = \frac{1}{S_{Y_i}^2} E\left(u_i^2\right) = \sigma^2$$

WLS矫正2（方差正比于自变量平方项）

矫正情形2：方差 σ_i^2 正比于 X_i^2

如果主模型存在异方差问题，且假设方差正比于 X_{2i}^2，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 X_{2i}^2$$



WLS矫正2（方差正比于自变量平方项）

对主模型两边同时除以 X_{2i}，得到：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \\
\frac{Y_i}{X_{2i}} &= \frac{\beta_1}{X_{2i}} + \beta_2 + \beta_3 \frac{X_{3i}}{X_{2i}} + \beta_4 \frac{X_{4i}}{X_{2i}} + \beta_5 \frac{X_{5i}}{X_{2i}} + \frac{e_i}{X_{2i}} \\
Y_i^* &= \beta_2^* X_{2i}^* + \beta_1^* + \beta_3^* X_{3i}^* \cdots + \beta_k^* X_{ki}^* + v_i
\end{aligned}
$$

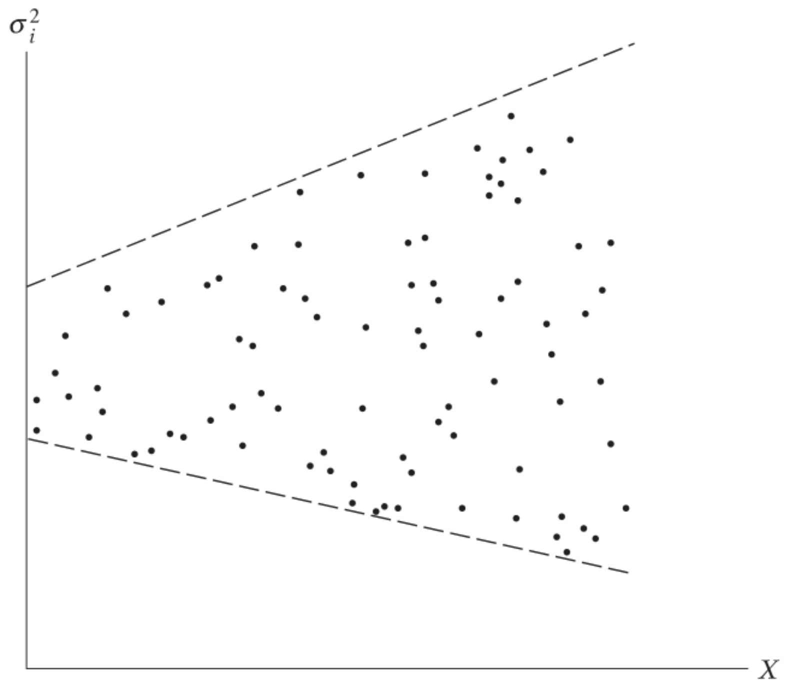
容易证明新模型将消除异方差性：

$$
E\left(v_i^2\right) = E\left(\frac{u_i}{X_{2i}}\right)^2 = \frac{1}{X_{2i}^2} E\left(u_i^2\right) = \sigma^2
$$

矫正情形3： 方差 σ_i^2 正比于 X_i

如果主模型存在异方差问题，且假设方差正比于 X_{2i}，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 X_{2i}$$



对主模型两边同时除以 $\sqrt{X_{2i}}$，得到：

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \\
\frac{Y_i}{\sqrt{X_{2i}}} &= \frac{\beta_1}{\sqrt{X_{2i}}} + \beta_2 \sqrt{X_{2i}} + \beta_3 \frac{X_{3i}}{\sqrt{X_{2i}}} + \cdots + \beta_5 \frac{X_{5i}}{\sqrt{X_{2i}}} + \frac{e_i}{\sqrt{X_{2i}}} \\
Y_i^* &= \beta_1^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + v_i
\end{aligned}
$$

容易证明新模型将消除异方差性：

$$
E\left(v_i^2\right) = E\left(\frac{u_i}{\sqrt{X_{2i}}}\right)^2 = \frac{1}{X_{2i}} E\left(u_i^2\right) = \sigma^2
$$

矫正情形4：方差 σ_i^2 正比于 $\hat{Y_i}^2$。如果主模型存在异方差问题，且假设方差正比于 $\hat{Y_i}^2$，则有：

$$var(u_i) = E(u_i^2) = \sigma_i^2 = \sigma^2 \hat{Y_i}^2$$

对主模型两边同时除以 $\hat{Y_i}$，得到：

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \\ \frac{Y_i}{\hat{Y_i}} &= \frac{\beta_1}{\hat{Y_i}} + \beta_2 \frac{X_{2i}}{\hat{Y_i}} + \cdots + \beta_k \frac{X_{ki}}{\hat{Y_i}} + \frac{e_i}{\hat{Y_i}} \\ Y_i^* &= \beta_1^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + v_i \end{aligned}$$

容易证明新模型将消除异方差性：

$$E\left(v_i^2\right) = E\left(\frac{u_i}{\hat{Y_i}}\right)^2 = \frac{1}{\hat{Y_i}^2} E\left(u_i^2\right) = \sigma^2$$

矫正情形5： 方差 σ_i^2 未知

如果主模型存在异方差问题，对主模型两边同时取对数 $ln()$通常能够减低异方差性问题。

$$
\begin{aligned}
Y_i &= \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \\
lnY_t &= \beta_1 + \beta_2 lnX_{2i} + \cdots + \beta_k lnX_{ki} + e_i
\end{aligned}
$$

WLS矫正（小结）

以上所有讨论的变换都是一种权宜之计。我们基本上是在猜测 σ_i^2 的情形。

> “幸福的家庭大抵类似，不幸的家庭各不相同!”——列夫托尔斯泰
>
> “同方差只有同一种结果（理想状态），异方差则可以异得千奇百怪（普遍现实）”

在所讨论的变换中哪一种能行之有效，要看问题的性质和异方差性的严重程度。这样做也会带来新的问题：

- 当我们超出双变量模型的范围时，我们也许不能预先知道应选择哪一个 X_{pi} 变量进行数据变换。
- 对数变换或开根号变换，当某些Y和X值为零或负数时便不适用。

使用White校正法矫正异方差的EViews报告如下：

Dependent Variable: Y
Method: Least Squares
Date: 04/02/19 Time: 23:28
Sample: 1 9
Included observations: 9
White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1990.668	484.0642	4.112404	0.0045
X	0.233148	0.049108	4.747687	0.0021

R-squared	0.438021	Mean dependent var	4161.767
Adjusted R-squared	0.357738	S.D. dependent var	420.6899
S.E. of regression	337.1460	Akaike info criterion	14.67204
Sum squared resid	795672.1	Schwarz criterion	14.71587
Log likelihood	-64.02418	Hannan-Quinn criter.	14.57746
F-statistic	5.455983	Durbin-Watson stat	0.616510
Prob(F-statistic)	0.052166	Wald F-statistic	22.54053
Prob(Wald F-statistic)	0.002089		

使用White校正法矫正异方差问题

作为对照，下面给出的是主模型的EViews报告：

Dependent Variable: Y
Method: Least Squares
Date: 04/02/19 Time: 15:32
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1990.668	936.2559	2.126200	0.0711
X	0.233148	0.099815	2.335805	0.0522

R-squared	0.438021	Mean dependent var	4161.767
Adjusted R-squared	0.357738	S.D. dependent var	420.6899
S.E. of regression	337.1460	Akaike info criterion	14.67204
Sum squared resid	795672.1	Schwarz criterion	14.71587
Log likelihood	-64.02418	Hannan-Quinn criter.	14.57746
F-statistic	5.455983	Durbin-Watson stat	0.616510
Prob(F-statistic)	0.052166		

3.2.6 案例展示

（异方差问题的诊断和矫正）

吸烟案例（数据表）

income	cigpric	educ	age	restaurn
20000	60.51	16	46	0
30000	57.88	16	40	0
30000	57.66	12	58	0
20000	57.88	13.5	30	0
20000	58.32	10	17	0
6500	59.34	6	86	0
20000	57.88	12	35	0

Showing 1 to 7 of 807 entries

Previous 1 2 3 4 5 … 116 Next

吸烟数量及其影响因素案例数据n=(807)

`cigs`作为因变量Y，表示吸烟数量（每天）；其余变量表示自变量X，其中：`income`表示年收入；`cigpric`表示每包香烟价格；`educ`表示读书年数；`age`表示年龄；`restaurn`表示是否禁止在餐馆公共场所吸烟。

首先构建如下**主回归模型**：

$$
\begin{aligned}
cigs = & +\hat{\beta}_1 && +\hat{\beta}_2 log(income) + \hat{\beta}_3 log(cigpric) + \hat{\beta}_4 educ \\
\text{(cont.)} & +\hat{\beta}_5 age + \hat{\beta}_6 I(age^2) && +\hat{\beta}_7 restaurn \quad +e_i
\end{aligned}
$$

主回归模型回归结果为：

$$
\begin{array}{lllll}
\widehat{cigs} = & -3.64 & +0.88log(income) & -0.75log(cigpric) & -0.50educ \\
\text{(t)} & (-0.1512) & (1.2095) & (-0.1301) & (-3.0016) \\
\text{(se)} & (24.0787) & (0.7278) & (5.7733) & (0.1671) \\
\text{(cont.)} & +0.77age & -0.01I(age^2) & -2.83restaurn & \\
\text{(t)} & (4.8132) & (-5.1765) & (-2.5410) & \\
\text{(se)} & (0.1601) & (0.0017) & (1.1118) & \\
\text{(fitness)} & R^2 = 0.0527; \bar{R}^2 = 0.0456 & & & \\
 & F^* = 7.42; \quad p = 0.0000 & & &
\end{array}
$$

吸烟案例（主模型EViews报告）

Dependent Variable: CIGS
Method: Least Squares
Date: 04/02/19 Time: 23:35
Sample: 1 807
Included observations: 807

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.639826	24.07866	-0.151164	0.8799
LOG(INCOME)	0.880268	0.727783	1.209519	0.2268
LOG(CIGPRIC)	-0.750862	5.773342	-0.130057	0.8966
EDUC	-0.501498	0.167077	-3.001596	0.0028
AGE	0.770694	0.160122	4.813155	0.0000
AGE^2	-0.009023	0.001743	-5.176494	0.0000
RESTAURN	-2.825085	1.111794	-2.541016	0.0112

R-squared	0.052737	Mean dependent var	8.686493
Adjusted R-squared	0.045632	S.D. dependent var	13.72152
S.E. of regression	13.40479	Akaike info criterion	8.037737
Sum squared resid	143750.7	Schwarz criterion	8.078448
Log likelihood	-3236.227	Hannan-Quinn criter.	8.053370
F-statistic	7.423062	Durbin-Watson stat	2.012825
Prob(F-statistic)	0.000000		

吸烟案例（残差数据）

得到主回归模型的残差序列 e_i、残差平方和序列 e_i^2（及其他变换数据）：

income	cigpric	educ	age	restaurn	e_i	e_i^2	$log(e_i^2)$	P_i
20000	60.51	16	46	0	-10.33	106.77	4.67	0.60
30000	57.88	16	40	0	-10.75	115.67	4.75	0.65
30000	57.66	12	58	0	-7.72	59.60	4.09	0.33
20000	57.88	13.5	30	0	-10.26	105.28	4.66	0.59
20000	58.32	10	17	0	-7.50	56.31	4.03	0.32
6500	59.34	6	86	0	2.44	5.95	1.78	0.03
20000	57.88	12	35	0	-11.93	142.42	4.96	0.80

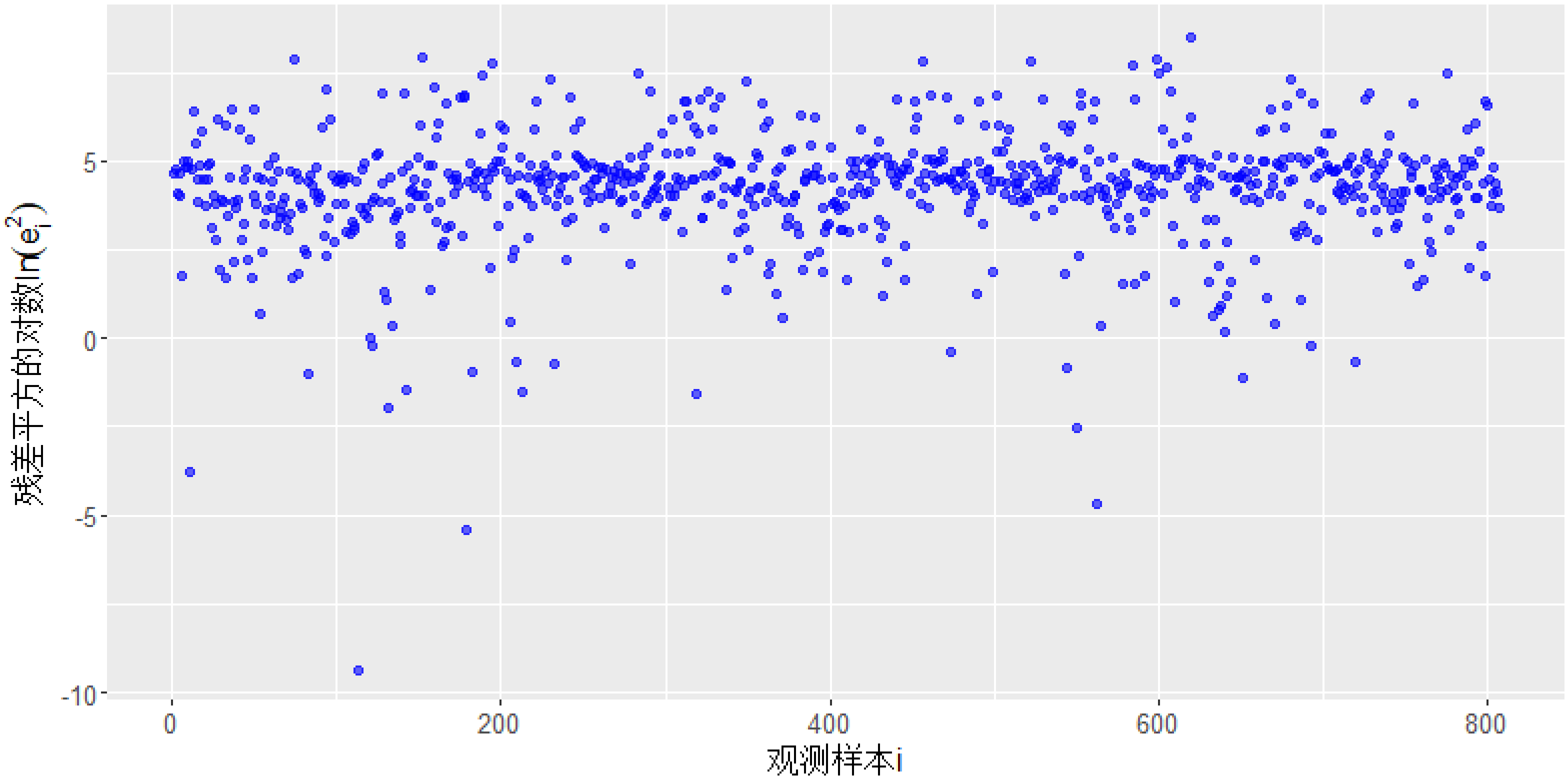
Showing 1 to 7 of 807 entries

Previous 1 2 3 4 5 … 116 Next

> **提示：** 后面BPJ检验中需要用到的数据 $P_i = \frac{\sum e_i^2}{\tilde{\sigma}^2}$; $\tilde{\sigma}^2 = \frac{\sum e_i^2}{n}$

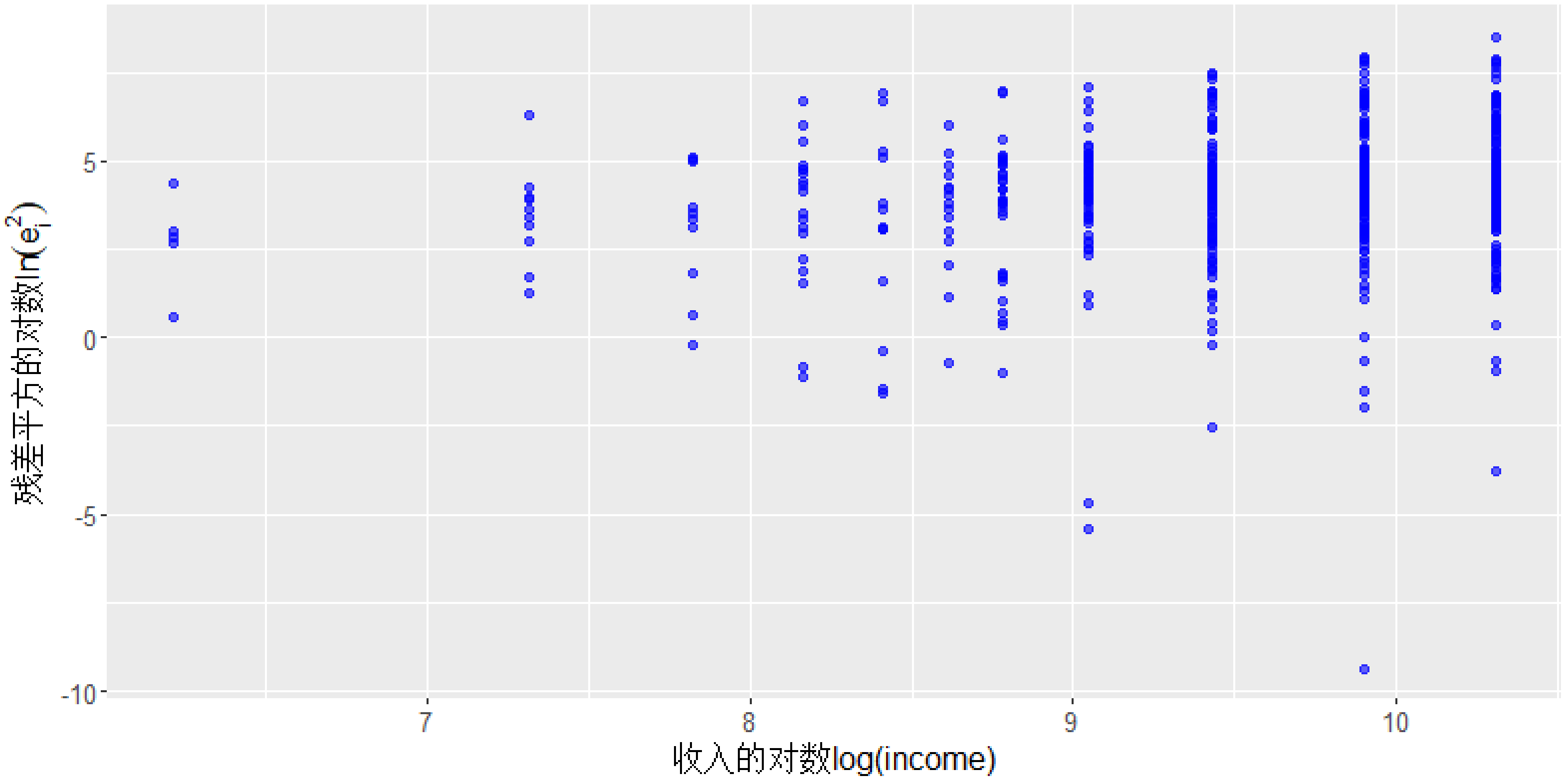
残差分布模式（按观测样本）

对数化残差平方序列 $log(e_i^2)$的**描点图**（dot plot）为：



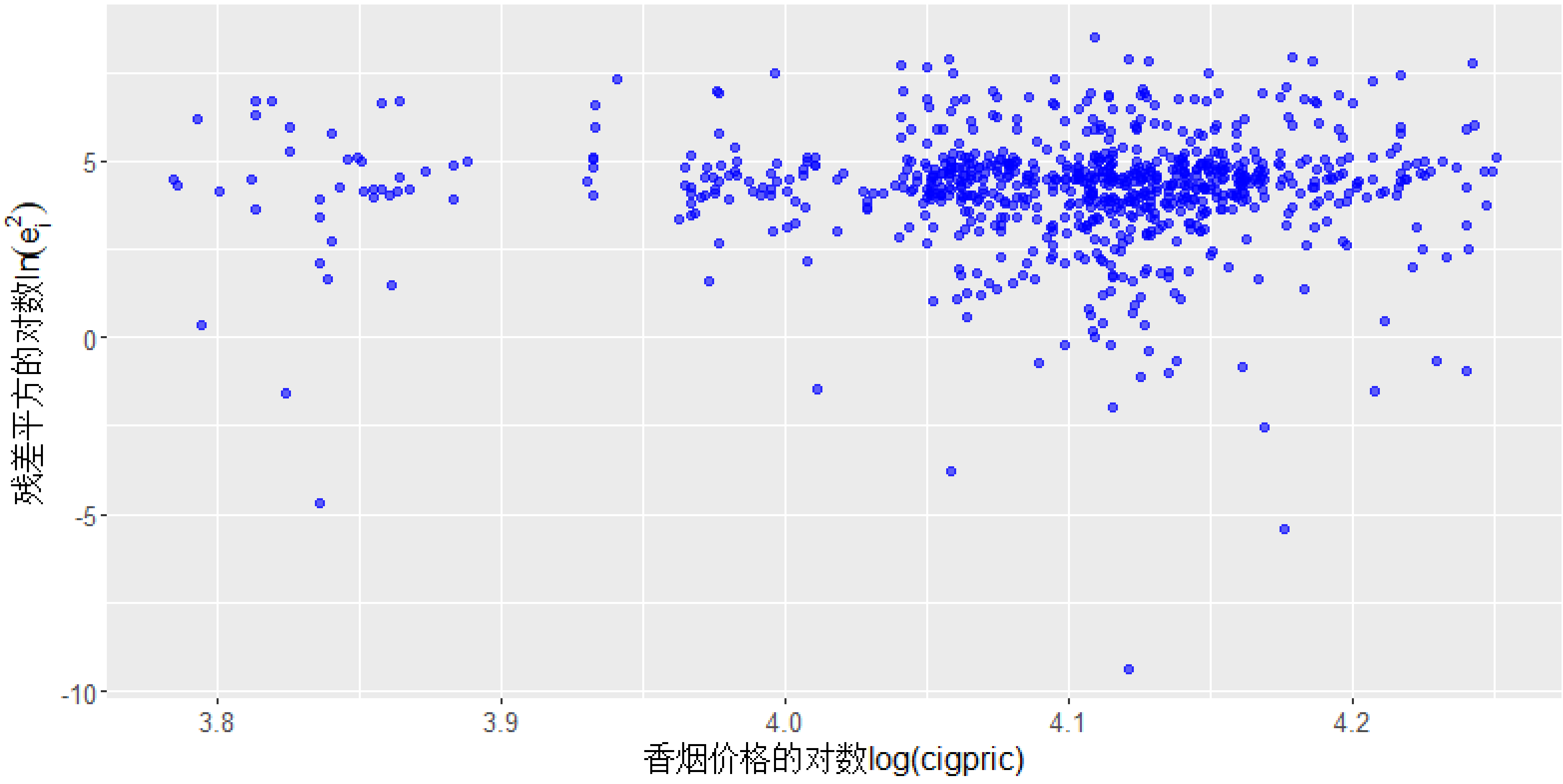
残差分布模式（对收入）

对数化残差平方序列 $log(e_i^2)$ 与**对数化**收入序列 $log(income)$ 的**散点图**（scatter plot）为：



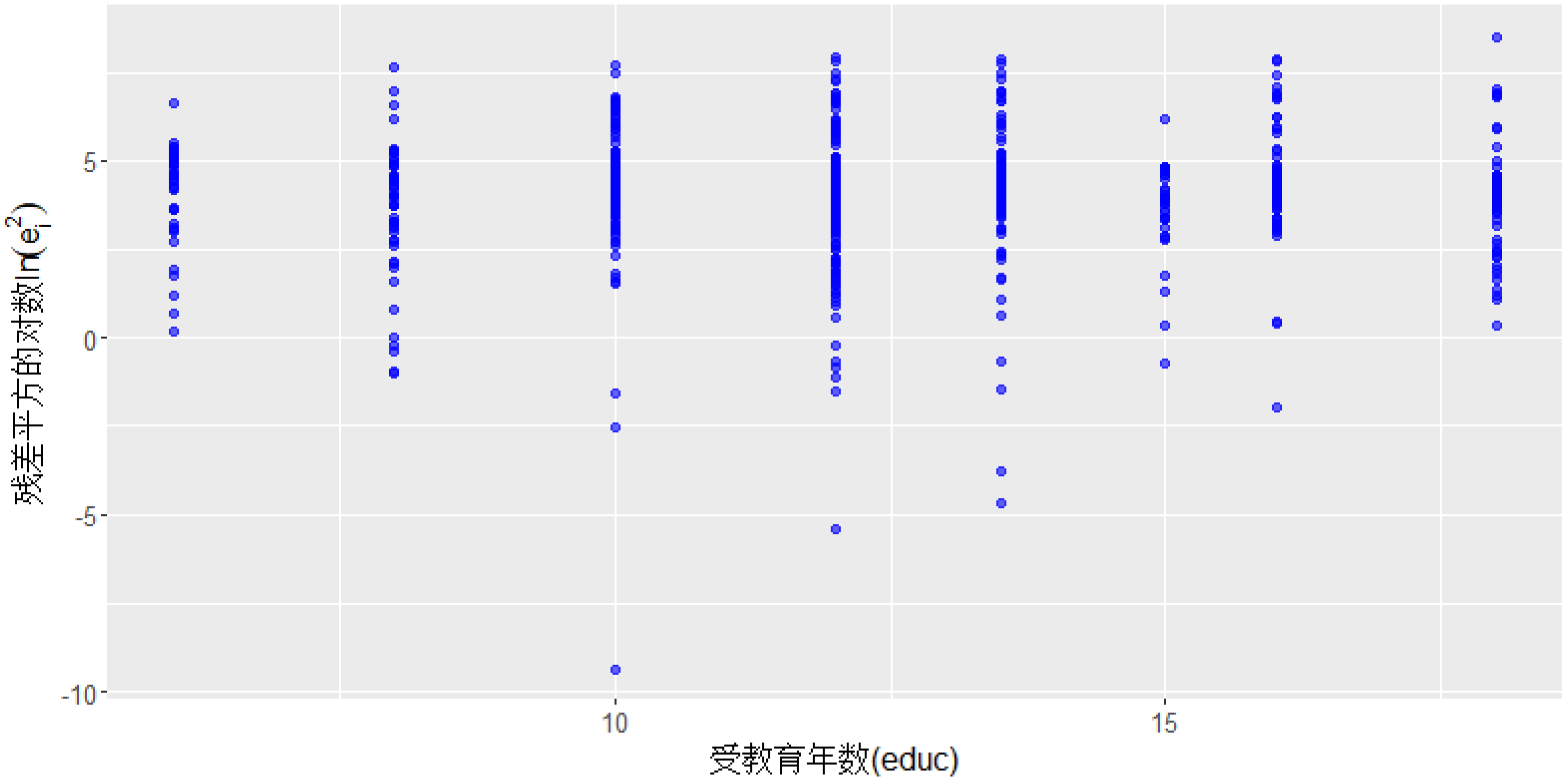
残差分布模式（对香烟价格）

对数化残差平方序列 $log(e_i^2)$ 与**对数化**香烟价格序列 $log(cigpric)$ 的**散点图**（scatter plot）为：



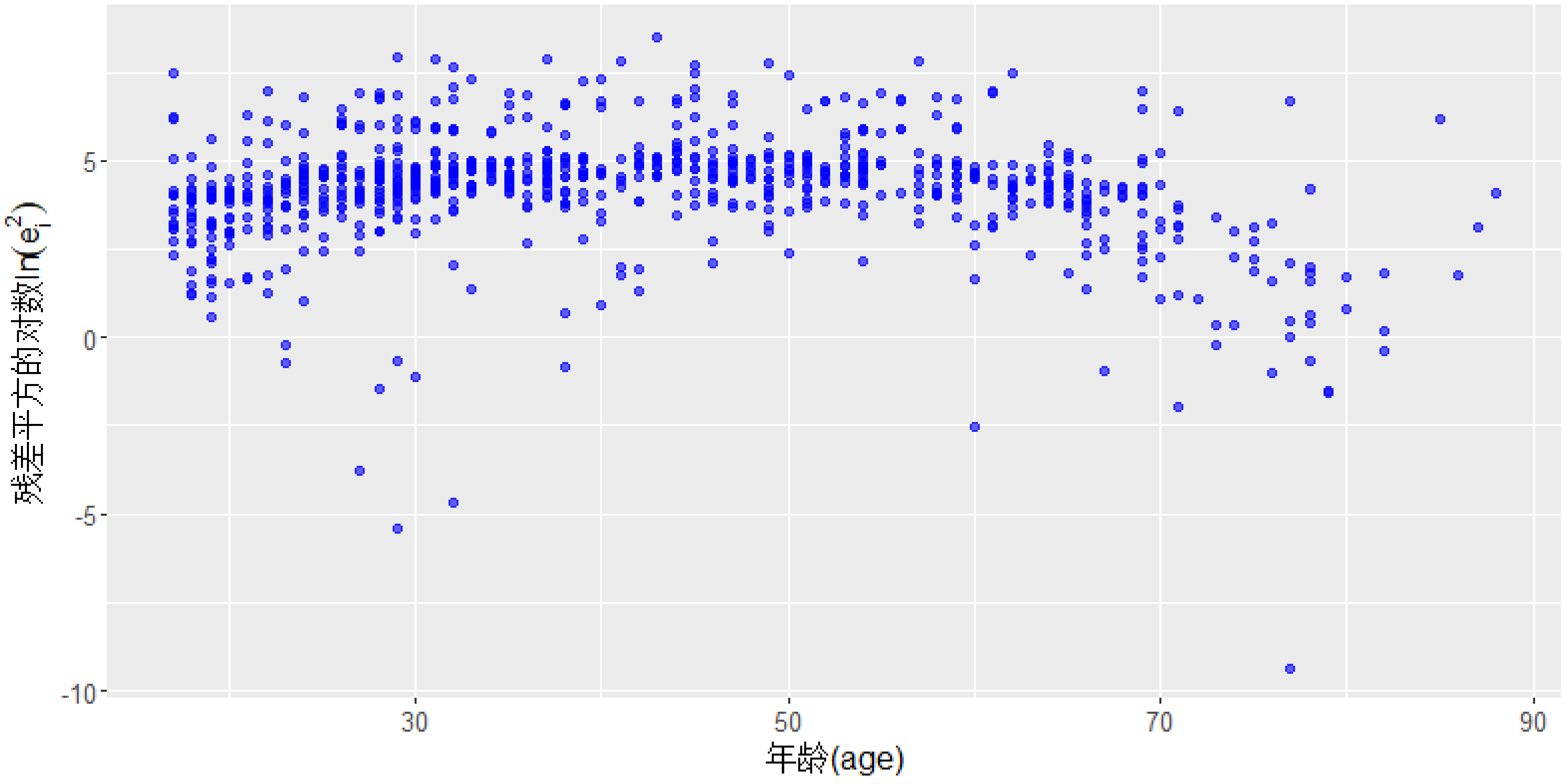
残差分布模式（对受教育年数）

对数化残差平方序列 $log(e_i^2)$与受教育年数序列 $educ$的**散点图**（scatter plot）为：



残差分布模式（对年龄）

对数化残差平方序列 $log(e_i^2)$与年龄序列 age的**散点图**（scatter plot）为：



帕克检验（关注收入变量）

为验证**主模型**是否存在异方差问题。我们可以构建如下的**帕克检验辅助方程1**

$$log(ei^2) = +\hat{\beta}_1 + \hat{\beta}_2 log(income) + e_i$$

帕克检验辅助方程的回归结果为：

$$\begin{array}{llll}
\widehat{log(ei^2)} = & -0.41 & & +0.48log(income) \\
(t) & (-0.5366) & & (6.0171) \\
(se) & (0.7700) & & (0.0793) \\
(fitness) & R^2 = 0.0430; \bar{R}^2 = 0.0419 \\
 & F^* = 36.21;\ p = 0.0000
\end{array}$$

从而可以得到**帕克检验**初步结论：F检验极显著，拒绝原假设 H_0，认为**主模型**为**异方差**。

为验证**主模型**是否存在异方差问题。我们可以构建如下的**BPG检验辅助方程**

$$
\begin{aligned}
Pi = \quad & + \hat{\beta}_1 \quad + \hat{\beta}_2 log(income) + \hat{\beta}_3 log(cigpric) + \hat{\beta}_4 educ \\
\text{(cont.)} & + \hat{\beta}_5 age + \hat{\beta}_6 I(age^2) \quad + \hat{\beta}_7 restaurn \quad + e_i
\end{aligned}
$$

BPG检验（LM检验）

BPG检验辅助方程的回归结果为：

$$
\begin{array}{lllll}
\widehat{Pi} = & -3.57 & +0.14log(income) & +0.34log(cigpric) & -0.01educ \\
\text{(t)} & (-0.9752) & (1.2493) & (0.3898) & (-0.5266) \\
\text{(se)} & (3.6630) & (0.1107) & (0.8783) & (0.0254) \\
\text{(cont.)} & +0.11age & -0.00I(age^2) & -0.40restaurn & \\
\text{(t)} & (4.4750) & (-4.5474) & (-2.3626) & \\
\text{(se)} & (0.0244) & (0.0003) & (0.1691) & \\
\text{(fitness)} & R^2 = 0.0400; \bar{R}^2 = 0.0328 & & & \\
 & F^* = 5.55; \quad p = 0.0000 & & &
\end{array}
$$

BPG检验辅助方程的回归平方和为 $ESS = 113.5691$；**LM统计量**（绝对值）为 $LM \equiv \chi^{2*} = ESS/2 = 56.7845$。给定显著性水平 $\alpha = 0.05$时卡方分布的理论查表值为 $\chi^2_{1-\alpha}(m-1) = \chi^2_{0.95}(1) = 12.59159$。从而可以得到**BPG检验**初步结论：卡方检验极显著，拒绝原假设 H_0，认为**主模型**存在**异方差性**。

怀特检验

怀特检验

我们可以构建如下**怀特检验辅助方程**

$$
\begin{aligned}
I(ei^2) = & + \hat{\beta}_1 && + \hat{\beta}_2 log(income) && + \hat{\beta}_3 I((log(income))^2) \\
(\text{cont.}) & + \hat{\beta}_4 log(cigpric) && + \hat{\beta}_5 I((log(cigpric))^2) && + \hat{\beta}_6 educ \\
(\text{cont.}) & + \hat{\beta}_7 age && + \hat{\beta}_8 I(age^2) && + \hat{\beta}_9 restaurn && + e_i
\end{aligned}
$$

怀特检验

进一步地，可以得到如下辅助回归的诊断结果：

$$
\begin{array}{llll}
\widehat{I(ei^2)} = & +19699.17 & +43.49log(income) & -1.04I((log(income))^2) \\
(t) & (1.0527) & (0.1572) & (-0.0683) \\
(se) & (18712.4188) & (276.6317) & (15.2518) \\
(cont.) & -10054.87log(cigpric) & +1251.96I((log(cigpric))^2) & -2.34educ \\
(t) & (-1.0865) & (1.0932) & (-0.5116) \\
(se) & (9254.3280) & (1145.2348) & (4.5649) \\
(cont.) & +19.51age & -0.22I(age^2) & -67.03restaurn \\
(t) & (4.4901) & (-4.5759) & (-2.2029) \\
(se) & (4.3442) & (0.0473) & (30.4296) \\
(fitness) & R^2 = 0.0414; & \bar{R}^2 = 0.0318 & \\
 & F^* = 4.31; & p = 0.0000 &
\end{array}
$$

怀特检验

怀特检验辅助方程的判定系数为 $R^2 = 0.0414$；样本卡方统计量（绝对值）为 $\chi^{2^*} = nR^2 = 33.4198$。给定显著性水平 $\alpha = 0.05$时卡方分布的理论查表值为 $\chi^2_{1-\alpha}(m-1) = \chi^2_{0.95}(1) = 15.50731$。从而可以得到**怀特检验**初步结论：卡方检验显著，应拒绝原假设 H_0，认为**主模型**存在**异方差问题**。

WLS矫正I（收入变量）

根据散点图，可以认为主模型方差正比于收入（income）的平方，因此加权最小二乘法（WLS）矫正模型如下：

$$
\begin{array}{llll}
I(\widehat{cigs/income}) = & +0.00 & +0.00I(log(cigpric/income)) & +0.07I(educ/income) \\
\text{(t)} & (5.4182) & (6.7093) & (2.1392) \\
\text{(se)} & (0.0006) & (0.0001) & (0.0331) \\
\text{(cont.)} & +0.00age & -0.00I(age^2) & -0.00restaurn \\
\text{(t)} & (3.6951) & (-4.0142) & (-2.3069) \\
\text{(se)} & (0.0000) & (0.0000) & (0.0001) \\
\text{(fitness)} & R^2 = 0.1486; \bar{R}^2 = 0.1432 & & \\
 & F^* = 27.95;\ p = 0.0000 & &
\end{array}
$$

我们可以构建如下的**BPG检验辅助方程**

$$
\begin{aligned}
Pi = & + \hat{\beta}_1 + \hat{\beta}_2 I(log(cigpric/income)) + \hat{\beta}_3 I(educ/income) \\
\text{(cont.)} & + \hat{\beta}_4 age + \hat{\beta}_5 I(age^2) + \hat{\beta}_6 restaurn + e_i
\end{aligned}
$$

BPG检验辅助方程的回归结果为：

$$
\begin{array}{llll}
\widehat{Pi} = & +3.77 & +0.95 I(log(cigpric/income)) & +1090.12 I(educ/income) \\
\text{(t)} & (1.4638) & (2.2419) & (8.0546) \\
\text{(se)} & (2.5730) & (0.4237) & (135.3412) \\
\text{(cont.)} & +0.08 age & -0.00 I(age^2) & -0.60 restaurn \\
\text{(t)} & (1.1710) & (-1.2426) & (-1.2390) \\
\text{(se)} & (0.0704) & (0.0008) & (0.4858) \\
\text{(fitness)} & R^2 = 0.1851; \bar{R}^2 = 0.1800 & & \\
 & F^* = 36.38;\ p = 0.0000 & &
\end{array}
$$

WLS矫正2（香烟价格变量）

根据散点图，也可以认为主模型方差正比于香烟价格（cigpric）的平方，因此加权最小二乘法（WLS）矫正模型如下：

$$
\begin{array}{lllll}
I(\widehat{cigs/cigpric}) = & -0.06 & +0.01I(log(income/cigpric)) & -0.39I(educ/cigpric) \\
\text{(t)} & (-0.7733) & (1.1926) & (-2.4633) \\
\text{(se)} & (0.0742) & (0.0123) & (0.1598) \\
\text{(cont.)} & +0.01age & -0.00I(age^2) & -0.06restaurn \\
\text{(t)} & (4.5097) & (-4.8636) & (-3.0092) \\
\text{(se)} & (0.0027) & (0.0000) & (0.0185) \\
\text{(fitness)} & R^2 = 0.0492; \bar{R}^2 = 0.0432 & & \\
 & F^* = 8.29; \quad p = 0.0000 & &
\end{array}
$$

我们可以构建如下的**BPG检验辅助方程**

$$
\begin{aligned}
Pi = & +\hat{\beta}_1 + \hat{\beta}_2 I(log(income/cigpric)) + \hat{\beta}_3 I(educ/cigpric) \\
\text{(cont.)} & + \hat{\beta}_4 age + \hat{\beta}_5 I(age^2) + \hat{\beta}_6 restaurn + e_i
\end{aligned}
$$

BPG检验辅助方程的回归结果为：

$$
\begin{array}{llll}
\widehat{Pi} = & -1.41 & +0.12 I(log(income/cigpric)) & +0.01 I(educ/cigpric) \\
\text{(t)} & (-2.1108) & (1.0434) & (0.0044) \\
\text{(se)} & (0.6675) & (0.1103) & (1.4379) \\
\text{(cont.)} & +0.10 age & -0.00 I(age^2) & -0.48 restaurn \\
\text{(t)} & (4.0588) & (-4.1341) & (-2.8722) \\
\text{(se)} & (0.0242) & (0.0003) & (0.1663) \\
\text{(fitness)} & R^2 = 0.0376; \bar{R}^2 = 0.0316 & & \\
 & F^* = 6.27; \quad p = 0.0000 & &
\end{array}
$$

对WLS矫正（香烟价格）模型进行BPG检验

BPG检验辅助方程的回归平方和为 $ESS = 117.8926$；**LM统计量**（绝对值）为 $LM \equiv \chi^{2^*} = ESS/2 = 58.9463$。给定显著性水平 $\alpha = 0.05$时卡方分布的理论查表值为 $\chi^2_{1-\alpha}(m-1) = \chi^2_{0.95}(4) = 9.487729$。从而可以得到**BPG检验**初步结论：卡方检验极显著，拒绝原假设 H_0，认为**WLS矫正模型**仍旧存在**异方差性**。

使用White校正法矫正异方差问题

使用White校正法矫正异方差的EViews报告如下：

使用White校正法矫正异方差问题

作为对照，下面给出的是主模型的EViews报告：

Dependent Variable: CIGS
Method: Least Squares
Date: 04/02/19 Time: 23:35
Sample: 1 807
Included observations: 807

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.639826	24.07866	-0.151164	0.8799
LOG(INCOME)	0.880268	0.727783	1.209519	0.2268
LOG(CIGPRIC)	-0.750862	5.773342	-0.130057	0.8966
EDUC	-0.501498	0.167077	-3.001596	0.0028
AGE	0.770694	0.160122	4.813155	0.0000
AGE^2	-0.009023	0.001743	-5.176494	0.0000
RESTAURN	-2.825085	1.111794	-2.541016	0.0112

R-squared	0.052737	Mean dependent var	8.686493
Adjusted R-squared	0.045632	S.D. dependent var	13.72152
S.E. of regression	13.40479	Akaike info criterion	8.037737
Sum squared resid	143750.7	Schwarz criterion	8.078448
Log likelihood	-3236.227	Hannan-Quinn criter.	8.053370
F-statistic	7.423062	Durbin-Watson stat	2.012825
Prob(F-statistic)	0.000000		

本章结束