计量经济学-作业8-异方差

😀 [redacted]

P200 习题2

2 回到第 5 章习题 9 关于跨国药品价格模型的分析，很显然，该研究是跨国家的（大到美国，小到卢森堡），所以，可以预期到异方差性是潜在问题。幸运的是，在原研究中，被解释变量是 P_i（P_i表示第 i 个国家与美国相比的相对药品价格水平，即第 i 个国家药品价格除以美国药品价格），所以，研究者并没有碰到通常会导致异方差性的规模差异（你知道这是为什么吗?）。

继续使用同样的数据，建立解释药品消费的估计方程为：

$$\widehat{CV}_i = -15.9 + 0.18N_i + 0.22P_i + 14.3IPC_i \qquad (10\text{-}21)$$

$$\begin{array}{cccc} & (0.05) & (0.09) & (6.39) \\ t = & 3.32 & 2.53 & 2.24 \end{array}$$

$$N = 32 \qquad \bar{R}^2 = 0.31$$

式中，CV_i代表第 i 个国家与美国的药品消费量之比；N_i代表第 i 个国家的人口除以美国人口的比值；IPC_i代表第 i 个国家是否鼓励价格竞争，为虚拟变量，如果是鼓励竞争则为 1，否则为 0。

a. 相对于 P 作为被解释变量，CV 作为被解释变量时，异方差性是否更容易出现？请解释理由。

b. 基于表 5-2 中数据（数据文件名为 DRUGS5），在 5% 的显著性水平下，运用 BP 检验和 White 检验，检验方程（10-21）是否存在异方差性。

c. 如果在问题 b 中得到的答案是存在异方差性，请计算方程（10-21）的 HC 标准差。

d. 同样，当存在异方差性时，采用双对数形式来重新估计方程（10-21）。

e. 同样，当存在异方差性时，重新定义方程（10-21）中的变量，并重新估计方程以消除异方差性。

f. 在以上三种异方差性补救方法中，哪一种最好？为什么？

g. 在第 5 章中，曾估计了一个方程，该方程以 P 为被解释变量，以 CVN（人均 CV）为解释变量。不同的是，在本例中 CV 是被解释变量，P 是解释变量。这样会违背哪个古典假定？请解释理由。

回答

a. 相对于用 `P` 作为被解释变量，用 `CV` 作为被解释变量时，**异方差更容易发生。**

   理由：因为药品的**总消费量（CV）**会随着国家的不同而发生巨大变化，而**药品价格(P)**的变化相对较小。

b. 检验过程如下：

i. BP检验

估计方程 $e_i^2 = a_0 + a_1 N_i + a_2 P_i + a_3 IPC_i + u_i$ (即异方差性来源于各解释变量)

建立假设 $H_0: a_1 = a_2 = a_3 = 0; H_A$: 假设H_0不成立

Equation: UNTITLED Workfile: DRUGS5::Drugs5\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic              4.825367    Prob. F(3,28)            0.0079
Obs*R-squared            10.90579    Prob. Chi-Square(3)      0.0122
Scaled explained SS      43.81744    Prob. Chi-Square(3)      0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/23/23   Time: 15:34
Sample: 1 32
Included observations: 32

       Variable       Coefficient   Std. Error  t-Statistic    Prob.

           C            -654.8450     300.1787    -2.181517     0.0377
           N             5.903529     2.051848     2.877176     0.0076
           P             7.415414     3.210487     2.309748     0.0285
         IPC             559.9163     239.8971     2.333985     0.0270

R-squared                0.340806    Mean dependent var      203.5942
Adjusted R-squared       0.270178    S.D. dependent var      670.1338
S.E. of regression       572.4929    Akaike info crite...    15.65435
Sum squared resid        9176946.    Schwarz criterion       15.83756
Log likelihood          -246.4695    Hannan-Quinn criter.    15.71508
F-statistic              4.825367    Durbin-Watson stat      1.607303
Prob(F-statistic)        0.007854
```

那么 $NR^2 == 10.90579$,其应该服从 $\chi^2(3) = 7.81$ 而11.075>7.81

因此拒绝原假设，认为方程存在异方差的问题。

ii. White检验

估计方程

$$
\begin{aligned}
e_i^2 =& a_0 + a_1 N_i + a_2 P_i + a_3 IPC_i \\
&+ a_4 N_i^2 + a_5 P_i^2 + a_6 IPC_i^2 + a_7 N_i P_i + a_8 P_i IPC_i + a_9 IPC_i N_i + u_i
\end{aligned}
$$

建立假设 $H_0: a_1 = a_2 = a_3 = \cdots = a_9 = 0; H_A$: 假设H_0不成立

Equation: UNTITLED Workfile: DRUGS5::Drugs5\

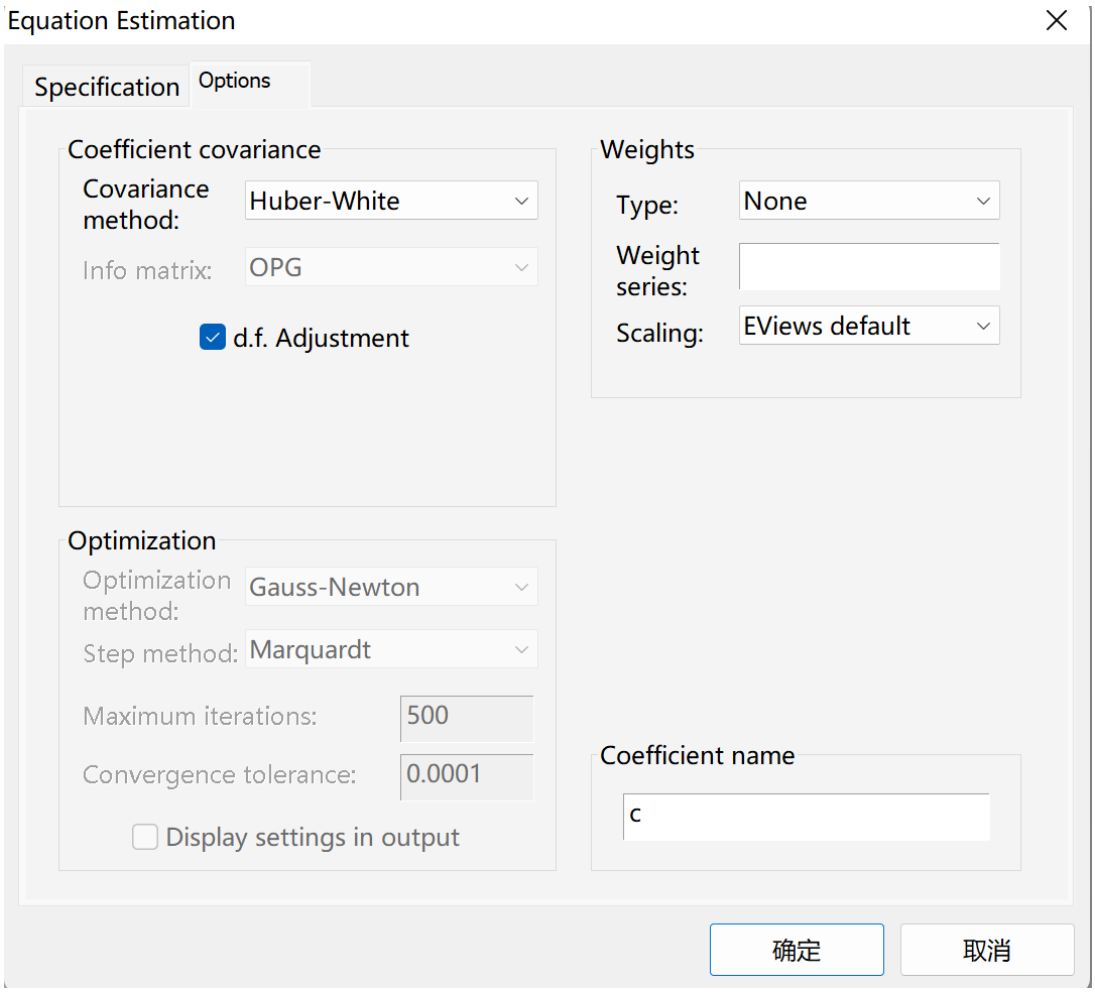
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic             24.34416    Prob. F(8,23)            0.0000
Obs*R-squared           28.62003    Prob. Chi-Square(8)      0.0004
Scaled explained SS     114.9900    Prob. Chi-Square(8)      0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/23/23   Time: 15:56
Sample: 1 32
Included observations: 32
Collinear test regressors dropped from specification

       Variable       Coefficient  Std. Error  t-Statistic   Prob.

           C            561.4637    398.8911     1.407561    0.1726
         N^2            0.089598    0.025209     3.554270    0.0017
         N*P            0.392586    0.092511     4.243650    0.0003
        N*IPC           24.97109    5.952052     4.195375    0.0003
           N           -36.92063    9.659131    -3.822356    0.0009
         P^2            0.024667    0.045727     0.539430    0.5948
        P*IPC           4.006431    8.139405     0.492227    0.6272
           P           -7.812420    8.975655    -0.870401    0.3931
        IPC^2          -571.9589    427.1334    -1.339064    0.1936

R-squared               0.894376    Mean dependent var      203.5942
Adjusted R-squared      0.857637    S.D. dependent var      670.1338
S.E. of regression      252.8483    Akaike info crite...    14.13571
Sum squared resid       1470442.    Schwarz criterion       14.54795
Log likelihood         -217.1714    Hannan-Quinn criter.    14.27236
F-statistic             24.34416    Durbin-Watson stat      2.435082
Prob(F-statistic)       0.000000
```

那么 $NR^2 == 28.62$,其应该服从 $\chi^2(9) = 16.9$ 而28.62>16.9

因此拒绝原假设，认为方程存在异方差的问题。

c. 用eviews软件进行怀特调整



调整的结果如下：

Equation: UNTITLED Workfile: DRUGS5::Drugs5\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Dependent Variable: CV
Method: Least Squares
Date: 10/23/23   Time: 16:04
Sample: 1 32
Included observations: 32
Huber-White-Hinkley (HC1) heteroskedasticity consistent standard
        errors and covariance
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-15.91351	11.59344	-1.372631	0.1808
N	0.181450	0.107206	1.692544	0.1016
P	0.216818	0.127345	1.702601	0.0997
IPC	14.33817	10.60658	1.351818	0.1873

R-squared	0.380607	Mean dependent var	8.842313
Adjusted R-squared	0.314243	S.D. dependent var	18.42017
S.E. of regression	15.25382	Akaike info crite...	8.404006
Sum squared resid	6515.015	Schwarz criterion	8.587223
Log likelihood	-130.4641	Hannan-Quinn criter.	8.464737
F-statistic	5.735179	Durbin-Watson stat	1.454755
Prob(F-statistic)	0.003440	Wald F-statistic	1.319168
Prob(Wald F-statis...	0.287856		

可以看到：N **的HC标准差是0.107；** P **的HC标准差是0.127；** IPC **的HC标准差是10.61。**

d. 采用**双对数**形式，那么解释变量和被解释变量都应用对数标识（IPC除外，因为IPC是虚拟变量，其含0值）

在Ewiews输入指令**LS log(CV) C log(N) log(P) IPC**，得到如下回归结果

Equation: UNTITLED Workfile: DRUGS5::Drugs5\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Dependent Variable: LOG(CV)
Method: Least Squares
Date: 10/23/23   Time: 16:09
Sample: 1 32
Included observations: 32
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.211842	1.971654	-4.164950	0.0003
LOG(N)	1.107918	0.139569	7.938160	0.0000
LOG(P)	1.461160	0.442280	3.303697	0.0026
IPC	0.875305	0.481676	1.817209	0.0799

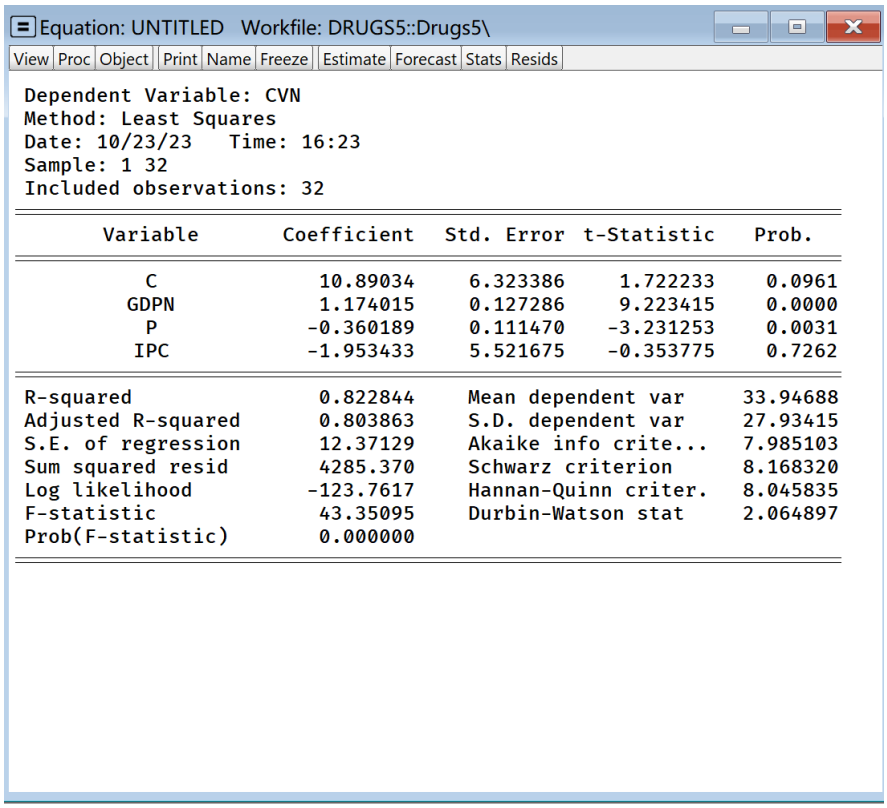
R-squared	0.733796	Mean dependent var	0.686299
Adjusted R-squared	0.705274	S.D. dependent var	2.048014
S.E. of regression	1.111840	Akaike info crite...	3.166378
Sum squared resid	34.61325	Schwarz criterion	3.349595
Log likelihood	-46.66205	Hannan-Quinn criter.	3.227109
F-statistic	25.72747	Durbin-Watson stat	1.293714
Prob(F-statistic)	0.000000		

即方程为

$$\widehat{\ln(CV_i)} = -8.21 + 1.11\ln(N_i) + 1.46\ln(P_i) + 0.88IPC_i$$
$$(0.14) \qquad (0.44) \qquad (0.48)$$
$$t = 7.94 \qquad 3.30 \qquad 1.82$$
$$N = 32 \qquad \overline{R^2} = 0.71$$

e. 重新定义变量，我们就需要消除规模的影响，我们以人均代替总数。CVN 代表人均，消费和GDP也非常相关，我们引入 $GDPN$ 代表人均GDP，

在Ewiews输入指令**LS CVN C GDPN P IPC**，得到如下回归结果



Equation: UNTITLED Workfile: DRUGS5::Drugs5\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Dependent Variable: CVN
Method: Least Squares
Date: 10/23/23   Time: 16:23
Sample: 1 32
Included observations: 32
```

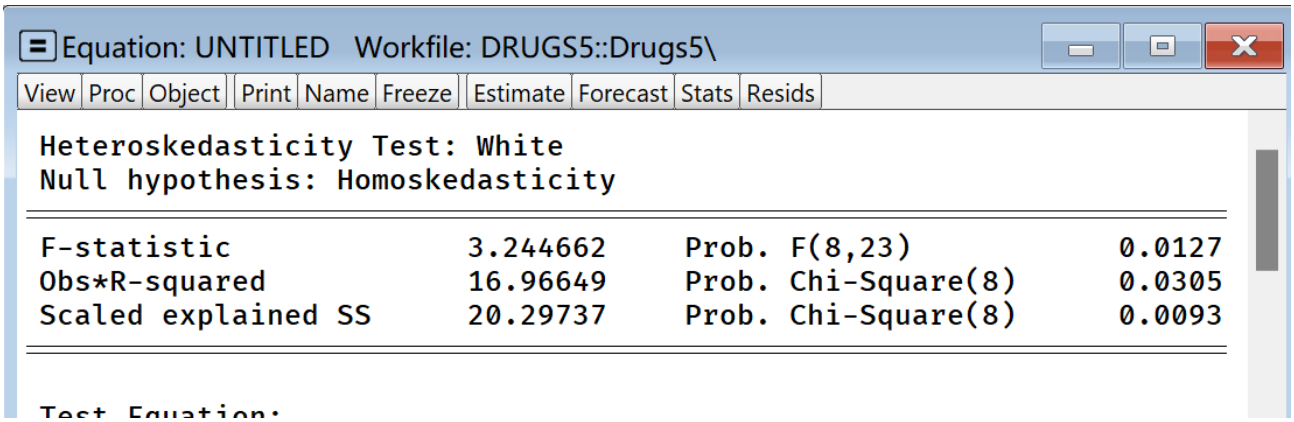
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.89034	6.323386	1.722233	0.0961
GDPN	1.174015	0.127286	9.223415	0.0000
P	-0.360189	0.111470	-3.231253	0.0031
IPC	-1.953433	5.521675	-0.353775	0.7262

R-squared	0.822844	Mean dependent var	33.94688
Adjusted R-squared	0.803863	S.D. dependent var	27.93415
S.E. of regression	12.37129	Akaike info crite...	7.985103
Sum squared resid	4285.370	Schwarz criterion	8.168320
Log likelihood	-123.7617	Hannan-Quinn criter.	8.045835
F-statistic	43.35095	Durbin-Watson stat	2.064897
Prob(F-statistic)	0.000000		

即方程为

$$\widehat{(CVN_i)} = 10.89 + 1.17GDPN_i - 0.36P_i - 1.95IPC_i$$
$$(0.13) \qquad (0.11) \qquad (5.52)$$
$$t = 7.94 \qquad 3.30 \qquad 1.82$$
$$N = 32 \qquad \overline{R^2} = 0.80$$

使用怀特检测，发现在5%显著性条件下，已经不存在异方差问题




```
Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic              3.244662    Prob. F(8,23)            0.0127
Obs*R-squared            16.96649    Prob. Chi-Square(8)      0.0305
Scaled explained SS      20.29737    Prob. Chi-Square(8)      0.0093

Test Equation:
```

f. 我认为采用**怀特调整（HC标准差）**最好，因为这样调整只改变了其标准差的大小，对其系数没有任何影响。保证了方程系数的不变性。且大多数学家都采用这种方法，比较主流。

g. 这个例子可能违背古典假设III（所有解释变量与误差项都不相关）。如果P是CV的函数，并且CV也是P的函数，那么，这就是一个联立方程系统，于是，误差项就不再独立于解释变量。

P200 习题4

4 A. 安多（A. Ando）和 F. 莫迪利亚尼（F. Modigliani）收集了关于非个体经济的房屋拥有者的收入和消费数据[⊖]。

收入范围（美元）	平均收入（美元）	平均消费（美元）	收入范围（美元）	平均收入（美元）	平均消费（美元）
0～999	556	2 760	5 000～5 999	5 538	5 320
1 000～1 999	1 622	1 930	6 000～7 499	6 585	6 250
2 000～2 999	2 664	2 740	7 500～9 999	8 582	7 460
3 000～3 999	3 587	3 515	10 000 以上	14 033	11 500
4 000～4 999	4 535	4 350			

a. 以平均消费为被解释变量，平均收入为解释变量，建立回归方程。
b. 针对以上方程中的残差，运用 BP 检验来检验异方差性，其中，显著水平为 5%。
c. 对同样的残差做 5%显著水平下的 White 检验。
d. 如果以上检验显示存在异方差性，应该如何解决？

回答

a. 建立的方程如下(CO_i 代表平均消费量 I 代表平均收入)

$$CO_i = \beta_0 + \beta_1 I_i$$

使用Eviews软件做回归，回归结果如下

Equation: UNTITLED Workfile: 新建 MICROSOFT EXCEL 工作表::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Dependent Variable: CO
Method: Least Squares
Date: 10/23/23   Time: 16:41
Sample: 1 9
Included observations: 9
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1273.164	291.9265	4.361250	0.0033
I	0.720442	0.044457	16.20553	0.0000

R-squared	0.974037	Mean dependent var	5091.667
Adjusted R-squared	0.970329	S.D. dependent var	3001.431
S.E. of regression	517.0087	Akaike info crite...	15.52713
Sum squared resid	1871086.	Schwarz criterion	15.57095
Log likelihood	-67.87207	Hannan-Quinn criter.	15.43255
F-statistic	262.6193	Durbin-Watson stat	1.469434
Prob(F-statistic)	0.000001		

即方程为——

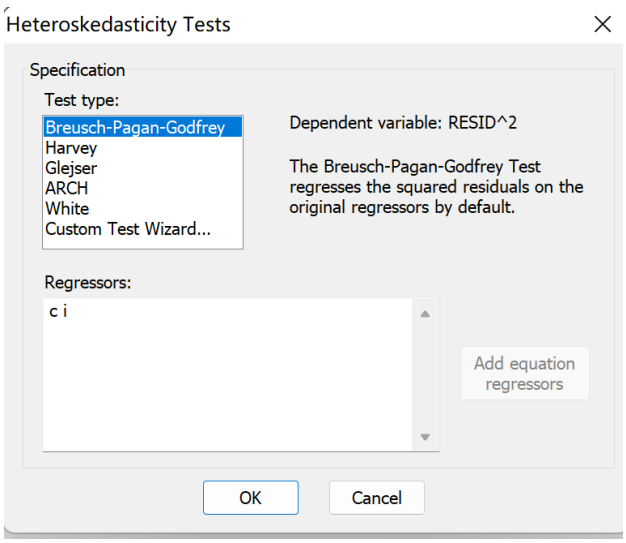
$$\widehat{CO_i} = 1273.164 + 0.72 I_i$$
$$(0.13)$$
$$t = 0.0444$$
$$N = 9 \qquad \overline{R^2} = 0.97$$

b. 进行BP检验

估计方程 $e_i^2 = a_0 + a_1 I_i + u_i$ (即异方差性来源于各解释变量)

建立假设 $H_0 : a_1 = 0; H_A$: 假设H_0不成立

在Eview中对估计的方程进行回归



检验步骤

Equation: UNTITLED Workfile: 新建 MICROSOFT EXCEL 工作表::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity
```

F-statistic	3.494083	Prob. F(1,7)	0.1038
Obs*R-squared	2.996617	Prob. Chi-Square(1)	0.0834
Scaled explained SS	2.637665	Prob. Chi-Square(1)	0.1044

```
Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/23/23   Time: 16:46
Sample: 1 9
Included observations: 9
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	487701.5	185451.1	2.629811	0.0339
I	-52.79081	28.24176	-1.869247	0.1038

R-squared	0.332957	Mean dependent var	207898.4
Adjusted R-squared	0.237666	S.D. dependent var	376167.4
S.E. of regression	328438.4	Akaike info crite...	28.43522
Sum squared resid	7.55E+11	Schwarz criterion	28.47904
Log likelihood	-125.9585	Hannan-Quinn criter.	28.34064
F-statistic	3.494083	Durbin-Watson stat	1.126196
Prob(F-statistic)	0.103793		

检验结果

那么 $NR^2 = 2.997$,其应该服从 $\chi^2(1) = 3.84$ 而2.997<3.84

因此不能拒绝原假设

c. **进行WHITE检验**

估计方程

$$e_i^2 = a_0 + a_1 I_i + a_2 I_i^2 + u_i$$

建立假设 $H_0: a_1 = a_2 = 0; H_A$: 假设H_0不成立

Equation: UNTITLED Workfile: 新建 MICROSOFT EXCEL 工作表::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic              6.637722    Prob. F(2,6)              0.0302
Obs*R-squared            6.198508    Prob. Chi-Square(2)       0.0451
Scaled explained SS      5.456015    Prob. Chi-Square(2)       0.0653

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/23/23   Time: 16:49
Sample: 1 9
Included observations: 9

       Variable       Coefficient   Std. Error  t-Statistic    Prob.

           C            903398.1     209578.4     4.310550     0.0050
          I^2           0.012385     0.004729     2.618691     0.0397
           I           -231.9756     71.52803    -3.243142     0.0176

R-squared               0.688723     Mean dependent var      207898.4
Adjusted R-squared      0.584964     S.D. dependent var      376167.4
S.E. of regression      242339.4     Akaike info crite...    27.89527
Sum squared resid       3.52E+11     Schwarz criterion       27.96101
Log likelihood         -122.5287     Hannan-Quinn criter.    27.75340
F-statistic             6.637722     Durbin-Watson stat      1.697743
Prob(F-statistic)       0.030161
```

那么 $NR^2 = 6.198$,其应该服从 $\chi^2(1) = 3.84$ 而6.198>3.84

因此拒绝原假设，认为方程存在异方差的问题。与BP检验结果背离。

d. 可以使用**HC标准差（怀特调整）**的方法对标准差进行调整。数据表中收入的范围不是不变的，所以所有变量应该是收入范围的中值。因此不同范围会使误差项产生不同的方差，因此也就带来了异方差的问题。通过怀特调整不会改变系数大小（也就是不同收入范围对支出的影响是不变的）

P200 习题6

6 R. 巴克林（R. Bucklin），R. 凯夫斯（R. Caves）和 A. 罗（A. Lo）估计了关于城市报纸年发行量的双对数模型。模型的估计结果为（括号内的数值为标准差）：[28]

$$\widehat{C_i} = -8.2 - 0.56P_i + 0.90I_i + 0.76Q_i + 0.27A_i + 0.08S_i - 0.77T_i$$
$$(0.58) \quad (0.14) \quad (0.21) \quad (0.14) \quad (0.05) \quad (0.27)$$
$$N = 50$$

式中，C_i代表第 i 种报纸的年发行量；P_i代表第 i 种报纸的加权平均价格；I_i代表第 i 种报纸所在城市的总可支配收入；Q_i代表第 i 种报纸的编辑相关人员数量；A_i代表第 i 种报纸的零售广告量；S_i代表第 i 种报纸所在地区的郊区日报的发行量；T_i代表第 i 种报纸所在地区的电视台数量。（以上所有变量都采用对数形式）。

a. 对每个斜率系数的符号都提出原假设，并做 t 检验。

b. 从理论上看，异方差性是否可能存在？请解释。

c. 根据你对于问题 a 和 b 的答案，这个方程可能存在哪些计量经济问题（从遗漏变量、变量不相关、错误的函数形式、多重共线性、序列相关性和异方差性中选择）？

d. 如果对方程的设定形式只做一项变换，应该怎样做？请解释原因。

回答

a. 在显著性水平5%的条件下分别对6个参数进行假设检验.其中，样本总数为50，可以求得自由度为 $n-k-1=50-6-1=43$,查表得单侧 $t_c=1.684$

- P :做出假设 $H_0:\beta \geq 0$ $H_A:\beta<0$,且预测参数为负。$t_1=\dfrac{-0.56}{0.58}=-0.97$，可知 $|-0.97|<1.684$ 即 $|t_1|<_c$，因此**不能拒绝** H_0
- I :做出假设 $H_0:\beta \leq 0$ $H_A:\beta>0$,且预测参数为正。$t_2=\dfrac{0.90}{0.14}=6.43$，可知 $|6.43|>1.684$ 即 $|t_2|>t_c$，且其预期符号和对立假设相同（均为正数），因此**可以拒绝** H_0
- Q :做出假设 $H_0:\beta \leq 0$ $H_A:\beta>0$,且预测参数为正。$t_3=\dfrac{0.76}{0.21}=3.62$，可知 $|3.62|>1.684$ 即 $|t_3|>t_c$，且其预期符号和对立假设相同（均为正数），因此**可以拒绝** H_0
- A :做出假设 $H_0:\beta \geq 0$ $H_A:\beta<0$,且预测参数为正。$t_4=\dfrac{0.27}{0.14}=1.93$，可知 $|1.93|>1.684$ 即 $|t_4|>t_c$,且其预期符号和对立假设相同（均为正数），因此**可以拒绝** H_0
- S :做出假设 $H_0:\beta \geq 0$ $H_A:\beta<0$,且预测参数为负 *（因为郊区发行越多，城区发行就会少）*。$t_4=\dfrac{0.08}{0.05}=1.6$，可知 $|1.6|<1.684$ 即 $|t_4|<t_c$,因此**不能拒绝** H_0
- T :做出假设 $H_0:\beta \leq 0$ $H_A:\beta>0$,且预测参数为正 *（因为电视台衡量了城市大小，而不是竞争对手）*。$t_6=\dfrac{-0.77}{0.27}=-2.85$，可知 $|-2.85|>1.684$ 即 $|t_5|<t_c$，但其预期符号和对立假设不相同，因此**不能拒绝** H_0

b. 这个模型**可能有异方差性**，异方差往往出现在规模一类的估计中，因为更大的城市会有更大的报纸发行量，而模型中的T也就衡量了这一大小，也导致了更大的误差项方差。

c. 我认为可能存在以下几个问题——

   i. 异方差：与第二小题的解释相同，更大的城市会有更大的报纸发行量

   ii. 存在遗漏变量的情况：方程中的符号有与预期不一致的情况，且从理论上来讲没有完全考虑如其他媒体的竞争因素等。编辑人数也并非一个很好的衡量报纸的质量的因素

   iii. 多重共线性：郊区发行量（S）和电视台数量（T）可能存在多重共线性，因为郊区发行量多可能的原因是城市规模小，这样的电视台数量也少。

d. **尝试改进方程的设定**。例如：用相互竞争的大都市的主流报纸的数量来替代S和T以减少多重线性，或者选用人们对报纸的评分代替编辑人数来衡量报纸的质量。