异方差性

随机误差项的方差不相同，（一）参数估计的无偏性仍然成立 参数估计的无偏性仅依赖于基本假定中的零均值 假定（即 ）。所以异方差的存在对无偏性 的成立没有影响。 （二）参数估计的方差不再是最小的 同方差假定是OLS估计方差最小的前提条件，所 以随机误差项是异方差时，将不能再保证最小二 乘估计的方差最小。由于异方差的影响，使得无法正确估计参数的标 准误差，导致参数估计的 t 统计量的值不能正确确 定，所以，如果仍用 t 统计量进行参数的显著性检 验将失去意义。

```
import scipy.stats as stats
import numpy as np
import statsmodels.api as sm
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn import model_selection

from patsy import dmatrices
```

```
df=pd.read_csv("chapter5.csv")
df
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	year	X1	X2	Y
0	1978.0	127.10	120.30	100.0
1	1979.0	155.90	142.10	102.0
2	1980.0	187.90	159.50	108.1
3	1981.0	221.00	184.00	110.7
4	1982.0	256.00	208.23	112.8

5	1983.0	258.40	231.12	114.5
6	1984.0	286.80	251.83	117.7
7	1985.0	315.07	276.25	128.1
8	1986.0	337.94	310.92	135.8
9	1987.0	369.46	348.32	145.7
10	1988.0	448.85	426.47	172.7
11	1989.0	494.07	473.59	203.4
12	1990.0	557.76	509.16	207.7
13	1991.0	590.21	552.39	213.7
14	1992.0	634.31	569.46	225.2
15	1993.0	698.27	647.43	254.9
16	1994.0	946.33	904.28	310.2
17	1995.0	1158.29	1061.15	356.1
18	1996.0	1453.42	1349.88	377.8
19	1997.0	1680.69	1440.48	380.8
20	1998.0	1789.17	1440.77	370.9
21	1999.0	1843.47	1426.06	359.8
22	2000.0	1903.60	1489.55	354.4
23	2001.0	1986.99	1497.52	351.6
24	2002.0	2107.66	1591.35	347.0
25	2003.0	2229.86	1747.02	346.7
26	2004.0	2580.28	2010.88	356.4
27	2005.0	2802.78	2274.17	359.4
28	2006.0	3002.38	2395.04	362.9
29	2007.0	3546.69	2747.27	376.7
30	2008.0	4121.21	3127.94	398.9
31	2009.0	4462.05	4141.40	400.5
32	2010.0	5139.52	3897.53	413.7

33	2011.0	6128.55	4103.92	435.2
34	2012.0	7001.43	5366.71	442.2
35	2013.0	7895.30	6126.80	449.7
36	2014.0	9348.00	8301.00	452.4
37	2015.0	10247.00	9251.00	453.3

```
y,X = dmatrices('Y ~ X1 + X2',data=df,return_type='dataframe')
model = sm.OLS(y,X)
fit = model.fit()
res = fit.resid
res_abs = abs(res)
DF1 = pd.DataFrame([])
DF1[['X1','X2']] = df[['X1','X2']]
DF1['res_abs'] = res_abs
print(DF1.corr())
#即可得到一个相关系数矩阵，通过变量与残差绝对值的相关系数可对模型是否存在异方差做出大致的判断
```

```
                X1        X2   res_abs
X1        1.000000  0.992367 -0.231247
X2        0.992367  1.000000 -0.206318
res_abs -0.231247 -0.206318  1.000000
```

White 检验

```
sm.stats.diagnostic.het_white(fit.resid, exog = fit.model.exog)
```

```
(5.9947452190068296,
 0.3067305673483387,
 1.1987521944186545,
 0.3320021205336836)
```

其中F检测的P-value为 0.332， 检验不通过 说明数据具有异方差性

修正方法：加权最小二乘法

```
model = sm.WLS(df.Y, df.loc[:, ['X1', 'X2']], weights=1. / (df.X1 ** 2)).fit()
print(model.summary())
```

```
                            WLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared (uncentered):                   0.805
Model:                            WLS   Adj. R-squared (uncentered):              0.794
Method:                 Least Squares   F-statistic:                              74.32
Date:                Wed, 15 Dec 2021   Prob (F-statistic):                    1.66e-13
Time:                        23:59:48   Log-Likelihood:                         -249.82
No. Observations:                  38   AIC:                                      503.6
Df Residuals:                      36   BIC:                                      506.9
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
X1            -0.9096      0.275     -3.308      0.002      -1.467      -0.352
X2             1.3926      0.322      4.331      0.000       0.741       2.045
==============================================================================
Omnibus:                        3.716   Durbin-Watson:                   0.290
Prob(Omnibus):                  0.156   Jarque-Bera (JB):                2.431
Skew:                           0.462   Prob(JB):                        0.297
Kurtosis:                       3.826   Cond. No.                         22.9
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```