异方差性: 误差方差不是常数会怎样?

黄光辉
hgh@cqu.edu.cn

目录

内容线索

- 本章的目标
- 异方差的性质
- 异方差对OLS估计的影响
- 广义最小二乘法
- OLS和GLS的区别
- 异方差的侦查方法
- 有异方差时的补救措施

本章的目标

我们寻求下述问题的解答:

1. 异方差的特征是什么?

本章的目标

我们寻求下述问题的解答:

1. 异方差的特征是什么?
2. 异方差对回归分析有什么后果?

本章的目标

我们寻求下述问题的解答:

1. 异方差的特征是什么?
2. 异方差对回归分析有什么后果?
3. 怎样去侦查异方差的存在性?

本章的目标

我们寻求下述问题的解答:

1. 异方差的特征是什么?
2. 异方差对回归分析有什么后果?
3. 怎样去侦查异方差的存在性?
4. 如果有异方差,怎样去补救?

内容线索

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。

反之，若扰动项方差不相同,则称为方差异性，也即异方差。

可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。公司盈利越多，可支付的红利越多.

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。公司盈利越多，可支付的红利越多.
4. 包含异常值的观测数据.

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。公司盈利越多，可支付的红利越多.
4. 包含异常值的观测数据.
5. 模型设定错误. 例如商品需求研究中,没有考虑其他替代商品和互补商品价格变化.

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。公司盈利越多，可支付的红利越多.
4. 包含异常值的观测数据.
5. 模型设定错误. 例如商品需求研究中,没有考虑其他替代商品和互补商品价格变化.
6. 模型中有一个或多个回归元取自分布不对称的总体.例如财富分配的分布，90%和10%的关系.

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。公司盈利越多，可支付的红利越多.
4. 包含异常值的观测数据.
5. 模型设定错误.例如商品需求研究中,没有考虑其他替代商品和互补商品价格变化.
6. 模型中有一个或多个回归元取自分布不对称的总体.例如财富分配的分布，90%和10%的关系.
7. 数据处理方法带来的情况。例如变形方法(一阶差分),或函数变换(取对数)。

可能出现异方差的情形

经典OLS假定：扰动项具有相同方差,称为同方差假定。
反之，若扰动项方差不相同,则称为方差异性，也即异方差。
可能出现方差异性的情形有:

1. 考虑家庭支出和收入的关系。收入越高,支出的变动越大.
2. 考虑边错边改的学习模型，例如打字练习中出错的个数。练习时间越长，出错个数越少.
3. 考虑公司红利支付政策。公司盈利越多，可支付的红利越多.
4. 包含异常值的观测数据.
5. 模型设定错误. 例如商品需求研究中,没有考虑其他替代商品和互补商品价格变化.
6. 模型中有一个或多个回归元取自分布不对称的总体.例如财富分配的分布，90%和10%的关系.
7. 数据处理方法带来的情况。例如变形方法(一阶差分),或函数变换(取对数)。
8. 异方差的重灾区: 横截面数据，忽略了成员规模差别.大型工厂和小型工厂,工人收入不同.

内容线索

- 本章的目标
- 异方差的性质
- 异方差对OLS估计的影响
- 广义最小二乘法
- OLS和GLS的区别
- 异方差的侦查方法
- 有异方差时的补救措施

异方差破坏了OLS估计的有效性

考虑回归模型:

$$Y = \beta_1 + \beta_2 X + U$$

参数$\widehat{\beta}_2$ 的方差:

满足同方差假定:	不满足同方差假定:
$Var\left(\widehat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2},$	$Var\left(\widehat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$

异方差破坏了OLS估计的有效性

考虑回归模型:

$$Y = \beta_1 + \beta_2 X + U$$

参数$\widehat{\beta}_2$ 的方差:

满足同方差假定:	不满足同方差假定:
$Var\left(\widehat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2}$,	$Var\left(\widehat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$

- 由于方差异性的存在,OLS估计不再是方差最小的线性估计.但是它依然是无偏估计和渐近估计.

异方差破坏了OLS估计的有效性

考虑回归模型:

$$Y = \beta_1 + \beta_2 X + U$$

参数$\widehat{\beta}_2$ 的方差:

满足同方差假定:	不满足同方差假定:
$Var\left(\widehat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2},$	$Var\left(\widehat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$

- 由于方差异性的存在,OLS估计不再是方差最小的线性估计.但是它依然是无偏估计和渐近估计.
- 如果异方差数据不进行数据转化,直接采用OLS得出的统计推断,例如t检验和F检验的结果带有误导.

异方差破坏了OLS估计的有效性

考虑回归模型:

$$Y = \beta_1 + \beta_2 X + U$$

参数$\widehat{\beta}_2$ 的方差:

满足同方差假定:	不满足同方差假定:
$Var\left(\widehat{\beta}_2\right) = \frac{\sigma^2}{\sum x_i^2}$,	$Var\left(\widehat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$

- 由于方差异性的存在,OLS估计不再是方差最小的线性估计.但是它依然是无偏估计和渐近估计.
- 如果异方差数据不进行数据转化,直接采用OLS得出的统计推断,例如t检验和F检验的结果带有误导.
- 转化数据常用广义最小二乘法.

内容线索

- 本章的目标
- 异方差的性质
- 异方差对OLS估计的影响
- 广义最小二乘法
- OLS和GLS的区别
- 异方差的侦查方法
- 有异方差时的补救措施

广义最小二乘法的思路

- 考虑线性模型:
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$

广义最小二乘法的思路

- 考虑线性模型:
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- 每次观测时,U_i 的方差设为σ_i^2.

广义最小二乘法的思路

- 考虑线性模型:
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- 每次观测时,U_i 的方差设为σ_i^2.
- 数据变形: $\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \frac{U_i}{\sigma_i}$

广义最小二乘法的思路

- 考虑线性模型:
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- 每次观测时,U_i 的方差设为σ_i^2.
- 数据变形: $\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \frac{U_i}{\sigma_i}$
- 新的回归记做: $Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + U_i^*$

广义最小二乘法的思路

- 考虑线性模型:
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- 每次观测时,U_i 的方差设为σ_i^2.
- 数据变形: $\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \frac{U_i}{\sigma_i}$
- 新的回归记做: $Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + U_i^*$
- 新模型满足同方差假定.

广义最小二乘法的思路

- 考虑线性模型:
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- 每次观测时,U_i 的方差设为σ_i^2.
- 数据变形: $\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \frac{U_i}{\sigma_i}$
- 新的回归记做: $Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + U_i^*$
- 新模型满足同方差假定.
- 每次观测所除的方差记为: $\omega_i = \frac{1}{\sigma_i}$,称为广义最小二乘法的权重.

广义最小二乘法的思路

- 考虑线性模型：
  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$
- 每次观测时,U_i 的方差设为σ_i^2.
- 数据变形: $\frac{Y_i}{\sigma_i} = \beta_1 \frac{X_{0i}}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \frac{U_i}{\sigma_i}$
- 新的回归记做: $Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + U_i^*$
- 新模型满足同方差假定.
- 每次观测所除的方差记为: $\omega_i = \frac{1}{\sigma_i}$,称为广义最小二乘法的权重.
- 对新模型进行OLS即可.

内容线索

- 本章的目标
- 异方差的性质
- 异方差对OLS估计的影响
- 广义最小二乘法
- OLS和GLS的区别
- 异方差的侦查方法
- 有异方差时的补救措施

两个最小二乘的目标函数

考虑线性模型：$Y = \beta_1 + \beta_2 X + U$,

两个最小二乘的目标函数

考虑线性模型：$Y = \beta_1 + \beta_2 X + U$,

OLS的目标函数:

$$\sum \widehat{u}_i^2 = \sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i \right)^2$$

两个最小二乘的目标函数

考虑线性模型：$Y = \beta_1 + \beta_2 X + U$,

OLS的目标函数:

$$\sum \widehat{u}_i^2 = \sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i\right)^2$$

GLS的目标函数:

$$\begin{aligned}
\sum \omega_i \widehat{u}_i^2 &= \sum \omega_i \left(Y_i - \widehat{\beta}_1^* - \widehat{\beta}_2^* X_i\right)^2 \\
\omega_i &= \frac{1}{\sigma_i^2}
\end{aligned}$$

两个最小二乘的目标函数

考虑线性模型：$Y = \beta_1 + \beta_2 X + U$,

$$
\begin{aligned}
&\text{OLS的目标函数:} \\
\sum \widehat{u}_i^2 &= \sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i\right)^2 \\
&\text{GLS的目标函数:} \\
\sum \omega_i \widehat{u}_i^2 &= \sum \omega_i \left(Y_i - \widehat{\beta}_1^* - \widehat{\beta}_2^* X_i\right)^2 \\
\omega_i &= \frac{1}{\sigma_i^2}
\end{aligned}
$$

OLS 和GLS 的区别:

1. OLS 是最小化残差平方和,而GLS是最小化加权的残差平方和.

两个最小二乘的目标函数

考虑线性模型：$Y = \beta_1 + \beta_2 X + U$,

OLS的目标函数:

$$\sum \widehat{u}_i^2 = \sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i \right)^2$$

GLS的目标函数:

$$\sum \omega_i \widehat{u}_i^2 = \sum \omega_i \left(Y_i - \widehat{\beta}_1^* - \widehat{\beta}_2^* X_i \right)^2$$

$$\omega_i = \frac{1}{\sigma_i^2}$$

OLS 和GLS 的区别:

1. OLS 是最小化残差平方和,而GLS是最小化加权的残差平方和.
2. OLS最优化时,每个扰动的权重是一样的; GLS最优化时,方差大的样本权重较小; 方差小的样本权重较大.

内容线索

- 本章的目标
- 异方差的性质
- 异方差对OLS估计的影响
- 广义最小二乘法
- OLS和GLS的区别
- 异方差的侦查方法
- 有异方差时的补救措施

异方差非正式的判别方法

可以从问题的性质和利用图形来判断是否存在异方差性.

1. 家庭预算研究中,残差方差随收入的增加而增加.

异方差非正式的判别方法

可以从问题的性质和利用图形来判断是否存在异方差性.

1. 家庭预算研究中,残差方差随收入的增加而增加.
2. 不均匀单元的横截面数据,例如投资与销售量的关系中,样本含有大,中,小型厂家.

异方差非正式的判别方法

可以从问题的性质和利用图形来判断是否存在异方差性.

1. 家庭预算研究中,残差方差随收入的增加而增加.
2. 不均匀单元的横截面数据,例如投资与销售量的关系中,样本含有大,中,小型厂家.
3. 从OLS残差$\widehat{u}_i^2$和$\widehat{Y}_i$的关系图,
   $\widehat{u}_i^2$和X_i的关系图,
   可定性判断是否存在异方差.
   多元回归多用后一种,一元回归多用前一种.

正式的统计检验方法

1. 帕克检验:

正式的统计检验方法

1. 帕克检验:

   $\log \widehat{u}_i^2 = \alpha + \beta \log X_i + v_i$

   $\widehat{u}_i$ 来自OLS. 两阶段线性回归,检验β的显著性.

正式的统计检验方法

1. 帕克检验:

   $\log \widehat{u}_i^2 = \alpha + \beta \log X_i + v_i$

   $\widehat{u}_i$ 来自OLS. 两阶段线性回归,检验β的显著性.

2. 葛莱泽检验:

   $|\widehat{u}_i| = \beta_1 + \beta_2 X_i + v_i$;

   $|\widehat{u}_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$;

   $|\widehat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$;

   $|\widehat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$.

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$
假定σ_i^2的正向关系为: $\sigma_i^2 = \sigma^2 X_i^2$, 也就是$X_i$越大, σ_i^2也越大.

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$
假定σ_i^2的正向关系为: $\sigma_i^2 = \sigma^2 X_i^2$, 也就是$X_i$越大, σ_i^2也越大.
戈德菲尔德-匡特检验步骤:

1. X的观测值从小到大排序;

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$
假定σ_i^2的正向关系为: $\sigma_i^2 = \sigma^2 X_i^2$, 也就是$X_i$越大, σ_i^2也越大.
戈德菲尔德-匡特检验步骤:

1. X的观测值从小到大排序;
2. 略去中间c个观测,前后数据自然分成两组,每组$(n-c)/2$;

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$
假定σ_i^2的正向关系为: $\sigma_i^2 = \sigma^2 X_i^2$, 也就是$X_i$越大, σ_i^2也越大.
戈德菲尔德-匡特检验步骤:

1. X的观测值从小到大排序;
2. 略去中间c个观测,前后数据自然分成两组,每组$(n-c)/2$;
3. 两组数据分别拟合,得到残差平方和RSS_1 和RSS_2,自由度$\frac{n-c}{2}-k$, k是包括截距项在内的参数个数;

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$
假定σ_i^2的正向关系为: $\sigma_i^2 = \sigma^2 X_i^2$, 也就是$X_i$越大, σ_i^2也越大.
戈德菲尔德-匡特检验步骤:

1. X的观测值从小到大排序;
2. 略去中间c个观测,前后数据自然分成两组,每组$(n-c)/2$;
3. 两组数据分别拟合,得到残差平方和RSS_1 和RSS_2,自由度$\frac{n-c}{2}-k$, k是包括截距项在内的参数个数;
4. 计算比例$\lambda = \frac{RSS_2/df}{RSS_1/df}$;

戈德菲尔德-匡特检验

这是一个广为流传的检验方差同回归变量之一有正向关系的检验.
考虑模型: $Y_i = \beta_1 + \beta_2 X_i + u_i$
假定σ_i^2的正向关系为: $\sigma_i^2 = \sigma^2 X_i^2$, 也就是$X_i$越大, σ_i^2也越大.
戈德菲尔德-匡特检验步骤:

1. X的观测值从小到大排序;
2. 略去中间c个观测,前后数据自然分成两组,每组$(n-c)/2$;
3. 两组数据分别拟合,得到残差平方和RSS_1 和RSS_2,自由度$\frac{n-c}{2}-k$, k是包括截距项在内的参数个数;
4. 计算比例$\lambda = \frac{RSS_2/df}{RSS_1/df}$;
5. 统计判断:

   如果扰动项u_i正态,并且同方差性为真实, 那么
   $\lambda \sim F\left(\frac{n-c}{2}-k, \frac{n-c}{2}-k\right)$.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.

布劳殊-培干- 戈弗雷可以避免这种局限性.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.

布劳殊-培干- 戈弗雷可以避免这种局限性.

布劳殊-培干-戈弗雷检验的步骤:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, 用OLS估计,得残差序列$\{\widehat{u}_i\}_{i=1}^n$.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.
布劳殊-培干- 戈弗雷可以避免这种局限性.
布劳殊-培干-戈弗雷检验的步骤:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, 用OLS估计,得残差序列$\{\widehat{u}_i\}_{i=1}^n$.
2. 计算σ^2的最大似然估计$\widehat{\sigma}^2 = \sum \widehat{u}_i^2/(n-k)$.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.
布劳殊-培干- 戈弗雷可以避免这种局限性.
布劳殊-培干-戈弗雷检验的步骤:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, 用OLS估计,得残差序列$\{\widehat{u}_i\}_{i=1}^n$.
2. 计算σ^2的最大似然估计$\widehat{\sigma}^2 = \sum \widehat{u}_i^2/(n-k)$.
3. 构造序列$p_i = \widehat{u}_i^2/\widehat{\sigma}^2$.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.

布劳殊-培干- 戈弗雷可以避免这种局限性.

布劳殊-培干-戈弗雷检验的步骤:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, 用OLS估计,得残差序列$\{\widehat{u}_i\}_{i=1}^n$.
2. 计算σ^2的最大似然估计$\widehat{\sigma}^2 = \sum \widehat{u}_i^2/(n-k)$.
3. 构造序列$p_i = \widehat{u}_i^2/\widehat{\sigma}^2$.
4. OLS 回归

   $p_i = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + v_i$
   诸Z_i可以是全部或部分X_i. 求得解释平方和ESS.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.
布劳殊-培干- 戈弗雷可以避免这种局限性.
布劳殊-培干-戈弗雷检验的步骤:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, 用OLS估计,得残差序列$\{\widehat{u}_i\}_{i=1}^n$.
2. 计算σ^2的最大似然估计$\widehat{\sigma}^2 = \sum \widehat{u}_i^2/(n-k)$.
3. 构造序列$p_i = \widehat{u}_i^2/\widehat{\sigma}^2$.
4. OLS 回归
   $p_i = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + v_i$
   诸Z_i可以是全部或部分X_i. 求得解释平方和ESS.
5. 定义$\Theta = \frac{1}{2} ESS$.

布劳殊-培干-戈弗雷检验(BGP)

戈德菲尔德-匡特检验需要确定c和X的某个观测从小到大的顺序（X多元时）.

布劳殊-培干- 戈弗雷可以避免这种局限性.

布劳殊-培干-戈弗雷检验的步骤:

1. $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, 用OLS估计,得残差序列$\{\widehat{u}_i\}_{i=1}^n$.
2. 计算σ^2的最大似然估计$\widehat{\sigma}^2 = \sum \widehat{u}_i^2/(n-k)$.
3. 构造序列$p_i = \widehat{u}_i^2/\widehat{\sigma}^2$.
4. OLS 回归
   $p_i = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_m Z_{mi} + v_i$
   诸Z_i可以是全部或部分X_i. 求得解释平方和ESS.
5. 定义$\Theta = \frac{1}{2} ESS$.
6. u_i为正态时,若扰动项同方差,样本容量足够大时有:
   $\Theta \sim \chi^2(m-1)$.

怀特的一般异方差性检验

戈德菲尔德-匡特检验需要选择X中的某个观测排序,
BGP易受偏离正态性的影响.
怀特检验不要求排序,也不依赖正态性假定,而且易于实施.

怀特的一般异方差性检验

戈德菲尔德-匡特检验需要选择X中的某个观测排序，
BGP易受偏离正态性的影响.
怀特检验不要求排序,也不依赖正态性假定,而且易于实施.
怀特检验的思路:(以三变量回归为例)

1. 估计$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$，得残差$\hat{u}_i$.

怀特的一般异方差性检验

戈德菲尔德-匡特检验需要选择X中的某个观测排序,
BGP易受偏离正态性的影响.
怀特检验不要求排序,也不依赖正态性假定,而且易于实施.
怀特检验的思路:(以三变量回归为例)

1. 估计$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$, 得残差$\widehat{u}_i$.
2. 估计辅助回归,求取R^2(还可引进回归元的高次方):
   $\widehat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$

怀特的一般异方差性检验

戈德菲尔德-匡特检验需要选择X中的某个观测排序,
BGP易受偏离正态性的影响.
怀特检验不要求排序,也不依赖正态性假定,而且易于实施.
怀特检验的思路:(以三变量回归为例)

1. 估计$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+u_i$, 得残差$\widehat{u}_i$.
2. 估计辅助回归,求取R^2(还可引进回归元的高次方):
   $\widehat{u}_i^2=\alpha_1+\alpha_2X_{2i}+\alpha_3X_{3i}+\alpha_4X_{2i}^2+\alpha_5X_{3i}^2+\alpha_6X_{2i}X_{3i}+v_i$
3. 假定无异方差,当n充分大时,有$nR^2\sim\chi^2(m)$, m为辅助回归中回归元(不含截距项)的个数.

怀特的一般异方差性检验

戈德菲尔德-匡特检验需要选择X中的某个观测排序,
BGP易受偏离正态性的影响.
怀特检验不要求排序,也不依赖正态性假定,而且易于实施.
怀特检验的思路:(以三变量回归为例)

1. 估计$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$, 得残差$\widehat{u}_i$.
2. 估计辅助回归,求取R^2(还可引进回归元的高次方):
   $\widehat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$
3. 假定无异方差,当n充分大时,有$nR^2 \sim \chi^2(m)$, m为辅助回归中回归元(不含截距项)的个数.
4. 统计推断:
   如果χ^2值超过临界值,结论是有异方差.(为什么?)

怀特的一般异方差性检验

戈德菲尔德-匡特检验需要选择X中的某个观测排序,
BGP易受偏离正态性的影响.
怀特检验不要求排序,也不依赖正态性假定,而且易于实施.
怀特检验的思路:(以三变量回归为例)

1. 估计$Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+u_i$, 得残差$\widehat{u}_i$.
2. 估计辅助回归,求取R^2(还可引进回归元的高次方):
   $\widehat{u}_i^2=\alpha_1+\alpha_2X_{2i}+\alpha_3X_{3i}+\alpha_4X_{2i}^2+\alpha_5X_{3i}^2+\alpha_6X_{2i}X_{3i}+v_i$
3. 假定无异方差,当n充分大时,有$nR^2\sim\chi^2(m)$, m为辅助回归中回归元(不含截距项)的个数.
4. 统计推断:
   如果χ^2值超过临界值,结论是有异方差.(为什么?)
   如果没有超过临界值,就说明个系数和零无差异,也就不存在异方差性.
5. 数据来源Table 11.3. 假想消费Y与收入X研究.

怀特检验解说

怀特检验的辅助回归含有所有回归元和他们的高次项,以及交叉项,需要估计的参数很多,对自由度的消耗很大.

怀特检验解说

怀特检验的辅助回归含有所有回归元和他们的高次项,以及交叉项,需要估计的参数很多,对自由度的消耗很大.

怀特检验的作用:

1. 怀特检验中若不含有交叉项,则是纯粹的异方差性检验;
2. 怀特检验若含有交叉项,则是异方差和模型设定误差的检验;

怀特检验解说

怀特检验的辅助回归含有所有回归元和他们的高次项,以及交叉项,需要估计的参数很多,对自由度的消耗很大.

怀特检验的作用:

1. 怀特检验中若不含有交叉项,则是纯粹的异方差性检验;
2. 怀特检验若含有交叉项,则是异方差和模型设定误差的检验; 若检验拒绝原假设, 有可能是由于异方差存在, 也有可能是模型设定误差.

寇因克-巴塞特(Koenker-Bassett)检验的思路:
若无异方差,则OLS残差和$\widehat{Y}^2$之间应该无线性关系.
KB检验的步骤:

1. 回归$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + U_i$

异方差性的其他检验: KB Test

寇因克-巴塞特(Koenker-Bassett)检验的思路:
若无异方差,则OLS残差和$\widehat{Y}^2$之间应该无线性关系.
KB检验的步骤:

1. 回归$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + U_i$
2. 估计辅助回归:
   $\widehat{u}_i^2 = \alpha_1 + \alpha_2 \left(\widehat{Y}_i\right)^2 + v_i$

异方差性的其他检验: KB Test

寇因克-巴塞特(Koenker-Bassett)检验的思路:
若无异方差,则OLS残差和$\widehat{Y}^2$之间应该无线性关系.
KB检验的步骤:

1. 回归$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + U_i$
2. 估计辅助回归:
   $$\widehat{u}_i^2 = \alpha_1 + \alpha_2 \left(\widehat{Y}_i\right)^2 + v_i$$
3. 统计推断:
   若$\alpha_2 = 0$,则不存在异方差性;
   若$\alpha_2 = 0$被拒绝,则存在异方差性.
   可以采用t检验或F检验.

异方差性的其他检验: KB Test

寇因克-巴塞特(Koenker-Bassett)检验的思路:
若无异方差,则OLS残差和$\widehat{Y}^2$之间应该无线性关系.
KB检验的步骤:

1. 回归$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + U_i$
2. 估计辅助回归:
   $\widehat{u}_i^2 = \alpha_1 + \alpha_2 \left(\widehat{Y}_i\right)^2 + v_i$
3. 统计推断:
   若$\alpha_2 = 0$,则不存在异方差性;
   若$\alpha_2 = 0$被拒绝,则存在异方差性.
   可以采用t检验或F检验.

KB检验的优点:
不依赖于误差项的分布是正态的假定.

内容线索

加权最小二乘法和怀特修正

有异方差时,采用加权最小二乘法和怀特方差修正.

1. 如果σ_i^2已知,采用加权最小二乘法,可以得到方差最小的线性无偏估计.

$$Y_i/\sigma_i = \beta_1^*/\sigma_i + \beta_2^* X_{2i}/\sigma_i + u_i/\sigma_i$$

加权最小二乘法和怀特修正

有异方差时,采用加权最小二乘法和怀特方差修正.

1. 如果σ_i^2已知,采用加权最小二乘法,可以得到方差最小的线性无偏估计.

$$Y_i/\sigma_i = \beta_1^*/\sigma_i + \beta_2^* X_{2i}/\sigma_i + u_i/\sigma_i$$

2. 如果σ_i^2未知,选择计算程序中的White Options,对OLS估计的方差作出适当的修正,使得由此得到的参数可以渐近有效,从而保证由此作出的统计推断是合适的.

加权最小二乘法和怀特修正

有异方差时,采用加权最小二乘法和怀特方差修正.

1. 如果σ_i^2已知,采用加权最小二乘法,可以得到方差最小的线性无偏估计.

$$Y_i/\sigma_i = \beta_1^*/\sigma_i + \beta_2^* X_{2i}/\sigma_i + u_i/\sigma_i$$

2. 如果σ_i^2未知,选择计算程序中的White Options,对OLS估计的方差作出适当的修正,使得由此得到的参数可以渐近有效,从而保证由此作出的统计推断是合适的.
3. 谨防对异方差性作出过度反应.若怀特修正后的方差和OLS估计的方差相差不大,那么可以不必进行修正.