四、异方差性

目录

异方差性及其产生的原因

- 异方差性及其产生的原因
  - 异方差性：$\mathrm{Var}(u_t) = \sigma_t^2 = \lambda_t^2 \sigma^2 \neq constant$
  - 产生原因：

    模型中遗漏了某些解释变量 x_t；

    模型函数形式的设定误差；

    样本数据的测量误差；

    截面数据中总体各单位的差异；

    随机因素的影响

异方差性的影响

- 一元线性回归模型下异方差性的影响 (OLS)
  - $\hat{b}_1, \hat{b}_0$ 是 b_1, b_0 的无偏、非有效估计量
  - $\hat{b}_1$ 不再有最小方差，$\mathrm{Var}(\hat{b}_1^*) = \mathrm{Var}(\hat{b}_1)\dfrac{\sum \lambda_t k_t^2}{\sum k_t^2} > \mathrm{Var}(\hat{b}_1)$
  - $s(\hat{b}_1)$ 一般会低估其真实方差，$s(\hat{b}_1^*) = s(\hat{b}_1)\sqrt{\dfrac{\sum \lambda_t k_t^2}{\sum k_t^2}} > s(\hat{b}_1)$，$t$ 检验的显著性会失去意义

异方差性的检验

- 异方差性的检验
  - 假设：$\mathrm{Var}(u_t) = E(u_t^2) \approx e_t^2$
  - $H_0 : u_t$ 为同方差性，$H_1 : u_t$ 为异方差性
  - 图示检验法 (相关图分析)：

    绘制 xOy 图与 xOe^2 图 (残差分布图)；

    若残差分布的离散程度有明显扩大的趋势，表明存在异方差性；

    只能粗略判断模型的异方差性

异方差性的检验

- 异方差性的检验
  - 戈德菲尔德-匡特检验 (G-Q 检验):

    将观测值按解释变量的大小顺序排列;

    将排在中间的约 $\frac{1}{4}$ 的观测值删除 (个数为 c), 将剩余观测值分成两部分, 每部分的观测值为 $\frac{n-c}{2}$ 个;

    对两部分分别进行回归分析, 计算残差 RSS_i;

    构造 $F = \frac{\mathrm{RSS}_2}{\mathrm{RSS}_1} \sim F\left(\frac{n-c}{2} - k - 1, \frac{n-c}{2} - k - 1\right)$;

    $F > F_\alpha\left(\frac{n-c}{2} - k - 1, \frac{n-c}{2} - k - 1\right)$ 或 $p < \alpha$, 拒绝 H_0, 认为存在异方差性; $F < F_\alpha$ 或 $p > \alpha$, 接受 H_0, 认为不存在异方差性;

    适用于样本容量较大、异方差性递增或递减的情况, 结果与 c 有关

异方差性的检验

- 异方差性的检验
  - 布罗斯-帕甘-戈费雷检验 (B-P-G 检验):

    辅助回归模型：$e_t^2 = a_0 + a_1 x_{1t} + \cdots + a_k x_{kt} + v_t$;

    F 统计量：$F = \dfrac{R^2/k}{(1-R^2)/(n-k-1)} \to F(n, n-k-1)$;

    LM 统计量：$\mathrm{LM} = nR^2 \to \chi^2(k)$;

    $F > F_\alpha$ 或 $\mathrm{LM} > \mathrm{LM}_\alpha$ 或 $p < \alpha$，拒绝 H_0，认为存在异方差性;

    $F < F_\alpha$ 或 $\mathrm{LM} < \mathrm{LM}_\alpha$ 或 $p > \alpha$，接受 H_0，认为不存在异方差性

异方差性的检验

- 异方差性的检验
  - 怀特检验 (White 检验)：

    二元线性回归的辅助回归模型：

    $e_t^2 = a_0 + a_1 x_{1t} + a_2 x_{2t} + a_3 x_{1t}^2 + a_4 x_{2t}^2 + a_5 x_{1t} x_{2t} + v_t$；

    一元线性回归的辅助回归模型：$e_t^2 = a_0 + a_1 x_t + a_2 x_t^2 + v_t$；

    统计量 nR^2 在二元下 $\to \chi^2(5)$，一元下 $\to \chi^2(2)$；

    $nR^2 > \chi_\alpha^2$ 或 $p < \alpha$，拒绝 H_0，认为存在异方差性；$nR^2 < \chi_\alpha^2$ 或 $p > \alpha$，接受 H_0，认为不存在异方差性

异方差性的检验

- 异方差性的检验
  - 戈里瑟检验 (Glejser 检验)：
    假定函数形式：$|e_t| = a_0 + a_1 x_t^h + v_t (h = \pm 1, \pm 2, \pm 1/2, \cdots)$；
    用 OLS 法计算 e_t，建立 $|e_t|$ 与 x_t^h 的回归方程，通过 p 值判断是否存在异方差性
  - 帕克检验 (Park 检验)：
    假定函数形式：$e_t^2 = a_0 x_1^{a_1} \mathrm{e}^{v_t} \iff \ln e_t^2 = \ln a_0 + a_1 \ln x_t + v_t$ 或
    $e_t^2 = a_0 + a_1 x_t + v_t$；
    用 OLS 法计算 e_t，建立 e_t 与 x_t^h 的回归方程，通过 p 值判断是否存在异方差性

异方差性的检验

- 异方差性的检验
  - ARCH 检验：

    用 OLS 法计算 e_t 与 $e_t^2, e_{t-1}^2, \cdots, e_{t-i}^2$；

    对辅助回归模型 $\hat{e}_t^2 = \hat{a}_0 + \hat{a}_1 e_{t-1}^2 + \cdots + \hat{a}_i e_{t-i}^2$ 进行回归分析，得到 R^2；

    计算统计量 $\mathrm{LM}(i) = (n-i)R^2 \sim \chi^2(i)$；

    $\mathrm{LM}(i) > \chi_\alpha^2(i)$ 或 $p < \alpha$，拒绝 H_0，认为存在异方差性；

    $\mathrm{LM}(i) < \chi_\alpha^2(i)$ 或 $p > \alpha$，接受 H_0，认为不存在异方差性
  - 检验异方差性的核心问题是判断随机误差项的方差与解释变量观测值之间的相关性

异方差性的解决方法

- 模型变换法 (一元)
  - 假设：$\mathrm{Var}(u_t) = \sigma_t^2 = \sigma^2 f(x_t)$
  - 令 $y_t^* = \dfrac{y_t}{\sqrt{f(x_t)}}, x_{1t} = \dfrac{1}{\sqrt{f(x_t)}}, x_{2t} = \dfrac{x_t}{\sqrt{f(x_t)}}, v_t = \dfrac{u_t}{\sqrt{f(x_t)}}$，则 $y_t^* = b_0 x_{1t} + b_1 x_{2t} + v_t$
  - 变换后，$\mathrm{Var}(v_t) = \sigma^2$
- 加权最小二乘法 (WLS)
  - $\min \sum w_t e_t^2$，取 $w_t = \sigma^2 f(x_t)$，重视小误差，轻视大误差
  - 步骤：

    用 OLS 法进行回归分析，求得 e_t；

    用 WLS 法进行回归分析，w_t 依据题意决定；

    用怀特检验 (White 检验) 判断是否还存在异方差性

异方差性的解决方法

- 模型的对数变换

取对数模型：$\ln y_t = b_0 + b_1 \ln x_t + u_t$，用 OLS 法对对数模型进行回归分析，对结果进行怀特检验 (White 检验) 判断是否还存在异方差性

- 广义最小二乘法
  - $E(\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}) = \sigma^2\boldsymbol{\Omega}$
  - $\hat{\boldsymbol{B}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{Y})$，是无偏、有效估计量
  - $\boldsymbol{\Omega} = [e_i e_j]_{n\times n}$
  - 若 $\boldsymbol{\Omega} = \boldsymbol{I}$，则 $\hat{\boldsymbol{B}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}$
  - 若 $\boldsymbol{\Omega}$ 为对角矩阵时，权矩阵 $\boldsymbol{P}$ 的元素 $p_{ij} = \begin{cases} 1/\sigma_i = 1/|e_i|, & i = j \\ 0, & i \neq j \end{cases}$

▶ 回归分析结果

同第二章：

常数和解释变量	参数估计值	参数标准误差	t统计量	双侧概率
C(b_0)	331.5264	57.16954	5.799003	0.0000
PI(b_1)	0.692812	0.006279	110.3337	0.0000
决定系数	0.997297	被解释变量均值		4662.514
调整的决定系数	0.997215	被解释变量标准差		4659.100
回归标准误差	245.8925	赤池信息准则		13.90311
残差平方和	1995283.	施瓦兹信息准则		13.99199
对数似然函数	−241.3044	汉南准则		13.93379
F统计量	12173.53	DW统计量		0.180221
F统计量的概率	0.000000			

代码输出结果分析

- 各种检验的输出结果分析

英文	含义	英文	含义
`Heterpskedasticity Test`	检验方法	`F-statistic`	回归模型的 F 统计量
`Obs*R-squared`	F 检验统计量	`Prob. Chi-Square(2)`	F 统计量对应的 p 值
`Prob. F(a, b)`	自由度为 a,b 的 F 分布临界值	`Scaled explained SS`	LM 统计量