

文章编号: 0529-6579 (1999) 02-0010-04

回归分析中异方差检验新方法^{*}

张华嘉, 舒 元

(中山大学岭南学院, 广东 广州 510275)

摘 要: 针对回归分析中异方差问题, 提出了新的 F -检验法和新的回归检验法, 并与 Goldfeld-Quandt 检验、White 检验和 Glejster 检验作了比较和讨论. 提出的新检验方法, 先对回归残差作线性变换, 然后再作检验, 与通用检验法相比, 检验的精确性大为提高.

关键词: 回归分析; 异方差; 假设检验

中图分类号: O 212 1 **文献标识码:** A

1 方法回顾

线性回归模型: $Y = X\beta + \epsilon$, 其中 $\epsilon \sim N(0, \Sigma)$, $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$. 异方差检验是指对 $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_N^2$ 进行假设检验. 在回归分析中, 除了残差图观察法外, 对异方差的检验方法可分为 2 类: F -检验法和回归检验法.

第 1 类是 F -检验法, 例如 Goldfeld-Quandt 检验^[1], 它利用 F -检验来检验方差是否与某变量 X_j 正(负)相关.

对某选定的变量 X_j 和正整数 d , 选择最大的 $(N-d)/2$ 个 X_j 为第 1 样本集 I_1 , 选择最小的 $(N-d)/2$ 个 X_j 为第 2 样本集, 使方程 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \dots + \beta_k X_{ki} + \epsilon_i$ ($i = 1 \sim N$) 分别在 I_1 和 I_2 上进行回归, 分别得到残差平方和 ESS_1 和 ESS_2 , 不妨设 $ESS_2 > ESS_1$, 则

$$\frac{ESS_2/\sigma_2^2}{ESS_1/\sigma_1^2} \sim F\left[\frac{N-d}{2}, \frac{N-d}{2}\right]$$

在零假设 $H_0: \sigma_1^2 = \sigma_2^2$ 成立下, $F = ESS_2/ESS_1 \sim F[(N-d)/2, (N-d)/2]$.

如果 $F > F_{\alpha}[(N-d)/2, (N-d)/2]$, 则 $\sigma_1^2 \neq \sigma_2^2$, 即方差变化与 X_j 有关, 如果 $F \leq F_{\alpha}[(N-d)/2, (N-d)/2]$, 则不能否定 $\sigma_1^2 = \sigma_2^2$, 即不能判定方差变化是否与 X_j 有关.

第 2 类是回归检验法, 例如 White^[2] 检验和 Glejster 检验^[3].

设 $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i$ ($i = 1 \sim N$)

对回归后的残差 e_i 的绝对值或平方再作回归,

White 检验: $e_i^2 = f(X_{1i}, X_{2i}, \dots, X_{ki}) + \epsilon_i$

Glejster 检验: $|e_i| = f(X_{1i}, X_{2i}, \dots, X_{ki}) + \epsilon_i$

^{*} 收稿日期: 1998-11-02

作者简介: 张华嘉, 男, 1964 年生, 讲师.

其中, f 是自己设计的函数形式, 如回归中某 X_i 显著, 则认为存在异方差.

本文利用 F -检验, 提出一种针对回归分析中异方差的新的检验方法, 与 Goldfeld-Quandt 检验相比, 不用对每个解释变量和潜在解释变量进行检验, 在统计量的样本选取上不受限制且有针对, 从而使检验效果更好. 另外, 指出 White 检验和 Glejster 检验在利用残差作回归时违反了普通最小二乘法的 2 个最基本的前提条件: 同方差, 样本不相关. 并提出异方差新的回归检验法, 它是对 White 检验和 Glejster 检验的改进.

本文提出 2 个新的检验方法, 与 Goldfeld-Quandt 检验、White 检验和 Glejster 检验相比, 具有先对回归残差作线性变换, 然后作 F -检验或回归分析的特点. 在目前条件下, 用计算机软件包 (例如 SAS) 对回归残差作线性变换是非常简单的事情. 结果使检验的精确性大为提高.

2 新的 F -检验法

设 $Y = X\beta + \epsilon$, $\epsilon \sim N(0, \Sigma)$, 其中 $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$

记 $P_X = X(X'X)^{-1}X'$ 为 X 的投影矩阵, 则有^[4] $P_X = A\Lambda A'$, 其中 A 为正交矩阵,

而 $\Lambda = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$

零假设 $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_N^2$. 记 $\sigma^2 = \sigma_1^2$.

定理 记 $A = (A_1, A_2)$, 其中 A_1 为 k 列, A_2 为 $N-k$ 列, 设 $Z = A_2'Y$. 则

1) $D(Z_j) = \sum a_{sj}^2 \sigma_s^2$, $\text{Cov}(Z_i, Z_j) = \sum a_{sj} a_{sj} \sigma_s^2$;

2) 在 H_0 成立的条件下有 $Z \sim N(0, \sigma^2 I_{N-k})$.

证明 由 P_X 为 Z 的投影矩阵

有 $P_X X = X$, 即 $(I - P_X)X = 0$, 又 $P_X = A\Lambda A'$,

故 $I - P_X = AA' - A\Lambda A' = A(I - \Lambda)A'$, 所以 $A(I - \Lambda)A'X = 0$,

从而 $(I - \Lambda)A'X = A'A(I - \Lambda)A'X = 0$,

因此 $(I - \Lambda)A'Y = (I - \Lambda)A'(X\beta + \epsilon) = (I - \Lambda)A'\epsilon$.

又 $I - \Lambda = \begin{pmatrix} 0 & 0 \\ 0 & I_{N-k} \end{pmatrix}$, 有 $(I - \Lambda)A' = A_2'$.

故 $A_2'Y = A_2'\epsilon$, 再由定义 $Z = A_2'Y$ 有 $Z = A_2'\epsilon$.

故 $D(Z) = D(A_2'\epsilon) = A_2'D(\epsilon)A_2 = A_2'\text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)A_2$.

所以 $D(Z_j) = \sum a_{sj}^2 \sigma_s^2$, $\text{Cov}(Z_i, Z_j) = \sum a_{sj} a_{sj} \sigma_s^2$.

在 H_0 成立条件下, $\epsilon \sim N(0, \sigma^2 I_N)$, 且 A_2 为列正交矩阵, 所以 $A_2'\epsilon \sim N(0, \sigma^2 I_{N-k})$.

再由 $Z = A_2'Y$ 即得 $Z \sim N(0, \sigma^2 I_{N-K})$ 证毕.

新的 F -检验方法:

设 $Z = (Z_1, Z_2, \dots, Z_{N-K})'$. 对 $|Z_i|$ 从大到小进行排序, 不妨设其次序为 $Z_1, Z_2, \dots, Z_{N-K}$, 设 m_1, m_2 为任意正整数且满足条件 $m_1 + m_2 \leq N - K$. 在 H_0 条件下, 由定理之

(2) 知 $Z_i \sim N(0, \sigma^2)$, ($i = 1 \sim N - K$), 所以

$$F_{m_1, m_2} = \frac{\sum_{i=1}^{N-k+m_2} Z_i^2 / m_1}{\sum_{i=N-m_1+1}^N Z_i^2 / m_2} \sim F(m_1, m_2)$$

如果存在 m_1, m_2 使 $F_{m_1, m_2} > F_\alpha(m_1, m_2)$, 则否定 H_0 , 即存在异方差. 如果对所有 m_1, m_2 有 $F_{m_1, m_2} \leq F_\alpha(m_1, m_2)$, 则不否定 H_0 , 即不能确定是否存在异方差.

本文方法与 Goldfeld-Quandt 检验进行比较, 有以下几个特点.

(1) Goldfeld-Quandt 检验是检验方差 σ_i^2 是否与某 X_i 有关, 为检验不存在异方差, 要用 G-Q 方法对每个解释变量和潜在解释变量进行检验, 但人们很难确定是否还有潜在解释变量. 本文 F -检验法检验是否存在异方差, 不需对每个解释变量和潜在解释变量进行检验.

(2) 本文 F -检验法在构造 F_{m_1, m_2} 时能随意确定 m_1, m_2 和按残差大小挑选样本, 如果其中某个 $F_{m_1, m_2} > F_\alpha(m_1, m_2)$, 则能确定异方差, 从而大大地提高了检验的效率.

(3) Goldfeld-Quandt 检验把样本集按某 X_i 的大小分为 2 个样本数一样的子集. 分别进行回归分析, 并从 2 个残差平方和之比的大小来推断方差 σ_i^2 是否与 X_i 有关, 但当 σ_i^2 在 X_i 的 2 端相差不大, 而在中间变化时, 例如 σ_i^2 与 X_i 的关系曲线呈 U 形, 这时 G-Q 检验就难以检验出 σ_i^2 与 X_i 的相关性; 而本文 F -检验法则可检出.

3 新的回归检验法

White 检验和 Glejster 检验都有 1 个共同思路, 就是看 ϵ_i 背后是否受到某个因素的影响, 即通过回归检验是否存在某个因素 X 使 $\text{Var}(\epsilon_i) = \sigma^2 f(X_i)$, 因为 $E(\epsilon_i^2) = \sigma_i^2$, 设 $\epsilon_i^2 = \sigma_i^2 + \hat{q}$, 在同方差假设下, $\sigma_i^2 = \sigma^2, i = 1 \sim N$, 有 $\epsilon_i^2 = \sigma^2 + \hat{q}$. 用原自变量的各种函数形式对 ϵ_i^2 进行回归, 一般找不到显著自变量. 反之如某自变量显著, 就会否定同方差假设.

由于不能计算 ϵ_i , White 检验和 Glejster 检验用残差 e_i 代替 ϵ_i , 但出现以下问题.

(1) 设 ϵ_i 同方差: $\sigma_i^2 = \sigma^2$, 因 $e = (I - P_X)Y$, 得 $D(e) = (I - P_X)\sigma^2$, 由于 $(I - P_X)$ 主对角线上数值不等, 使 e_i 异方差, 也就是说, 即使 ϵ_i 同方差, e_i^2 也可能受到某自变量的显著影响.

(2) $D(e) = (I - P_X)\sigma^2$ 可知, 只有在 X 列正交时, P_X 才是对角阵, 从而 $I - P_X$ 为对角阵, 否则存在相关的 e_i, e_j . 另外, 矩阵 $(I - P_X)$ 的秩为 $N - K$, 由 $e = (I - P_X)Y$ 得知, N 个残差样本中只有 $N - K$ 个独立样本. 但普通最小二乘法前提假设是样本不相关, 否则会影响回归的各种统计检验, 因此, 一般不能用 e 来代替 ϵ 进行回归.

新的回归检验法:

设 $\text{Var}(\epsilon_j) = f(X_{1j}, X_{2j}, \dots, X_{kj})$, 由定理(1)得 $D(Z_j) = \sum a_{sj}^2 \sigma_j^2 = \sum a_{sj}^2 f(X_{1s}, X_{2s}, \dots, X_{ks})$, 因为 $Z_j \sim N(0, D(Z_j))$, 所以 $ED_j^2 = D(Z_j)$.

设 $Z_j^2 = D(Z_j) + \hat{q}$, $\sigma \sim N(0, \sigma^2 I_{N-K})$, 则 $Z_j^2 = \sum a_{sj}^2 f(X_{1s}, X_{2s}, \dots, X_{ks}) + \hat{q}$, 对其进行回归分析.

在 ϵ 同方差条件下, 由定理(2)得 $Z \sim N(0, \sigma^2 I_{N-K})$, 回归对各解释变量应该不显著, 如果显著非零, 就能否定同方差假设.

例 $f(X_{1j}, \dots, X_{kj}) = \alpha + \beta X_{1j} + \gamma X_{1j}^2$, 由定理(1)得

$$D(Z_j) = \sum a_{sj}^2 \sigma_s^2 = \alpha \sum a_{sj}^2 + \beta \sum a_{sj}^2 X_{1s} + \gamma \sum a_{sj}^2 X_{1s}^2.$$

因为 $Z_j \sim N(0, D(Z_j))$, 所以 $EZ_j^2 = D(Z_j)$.

设 $Z_j^2 = D(Z_j) + \hat{q}$, $\delta \sim N(0, \sigma^2 I_{N-K})$, 则 $Z_j^2 = \sum \alpha + \beta \sum a_{sj}^2 X_{1j} + \gamma \sum a_{sj}^2 X_{1j}^2 + \hat{q}$.

对其进行回归分析.

在 ϵ 同方差条件下, 由定理(2)得 $Z \sim N(0, \sigma^2 I_{N-K})$, β, γ 应该不显著, 如果 β, γ 在回归中显著非零, 就能否定同方差假设. 类似地, 可以把 $f(X_{1j}, X_{2j}, \dots, X_{kj})$ 推广到多元多项式情形, 进一步可推广到因变量与参数之间的关系为线性关系的函数情形.

在用 $\text{Var}(\epsilon_i) = f(X_{1j}, X_{2j}, \dots, X_{kj})$ 检验出存在异方差后, 要对 σ_j^2 进行估计, 并用加权最小二乘法修正. 严格地说, 即使估计结果很好, 上述检验式 $\sigma_j^2 = f(X_{1j}, X_{2j}, \dots, X_{kj})$ 也不一定能为 σ_j^2 的估计, 因为在否定了同方差时, 由定理(1)知 $\text{Cov}(Z_i, Z_j) = \sum a_{sj} a_{sj} \sigma_s^2$ 可能不为零, 即 Z_i, Z_j 可能相关. 上述回归与普通最小二乘法样本不相关假设矛盾, 因此回归的统计检验可能因序列相关出错. 但在实际应用中, 可以把严格性放松, 认为 $\sigma_i^2 = f(X_{1j}, X_{2j}, \dots, X_{kj})$, 从而用加权最小二乘法修正.

新的回归检验法对 White 检验和 Glejster 检验进行了改进, 先对残差作变换 $Z = A_2' Y$, 并对自变量作变换后再进行回归分析. 虽然样本数从原来 N 变为 $N-k$, 但在同方差假设条件下, Z 满足普通最小二乘法的 2 个最基本的条件: 同方差, 样本不相关. 而 White 检验和 Glejster 检验所使用的残差 e_i 不能满足上述条件.

参考文献:

- [1] GOLDFELD S M, QUANDT R E. Some test for homoscedasticity [J]. J Ame Sta Soc, 1965, 60: 539 ~ 547.
- [2] WHITE H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity [J]. Econometrica, 1980, 48: 817 ~ 838.
- [3] 刘振亚. 计量经济学教程 [M]. 北京: 中国人民大学出版社, 1997. 190.
- [4] 杨维权, 刘兰亭, 林鸿洲. 多元统计分析 [M]. 北京: 高等教育出版社, 1991. 285 ~ 287.

New Methods to Test of Heteroskedasticity

ZHANG Hua-jia^{*}, SHU Yuan

Abstract: New methods with F -test and with regression are put forward for heteroskedasticity. They are compared with Goldfeld-Quandt test, White test and Glejster test. The new tests apply linear transform to the residuals first, then test them, resulting in higher accuracy, compared with the former ones.

Keywords: regression; heteroskedasticity; test

^{*} Lingnan College, Zhongshan University, Guangzhou 510275, China