

异方差的游程检验^{*}

兰嘉庆, 余宛玲
(西南财经大学, 四川 成都 610074)

摘要: 异方差现有常规检验方法的大多是参数检验的方法, 需要参数服从一定的分布。但实际研究中往往有很多参数不服从假设的分布, 针对这一以往方法的缺陷, 提出了异方差的游程检验方法。该方法能克服参数检验需要参数分布的缺陷, 而且也能用于小样本。先对游程检验做了简单的概述, 介绍了游程检验的基本原理, 然后就叙述了游程检验对异方差检验的应用。

关键词: 异方差; 游程检验; 游程

中图分类号: O212.1 **文献标识码:** A **文章编号:** 0529-6579 (2004) S1-0009-03

经济计量模型研究中一个不可避免的问题是所谓“异方差性”问题, 或曰“方差扩大化”问题。产生异方差后还用最小二乘法去估计参数, 会导致以下的严重后果: ①参数估计值不再具有最小方差的特性; ②解释变量的显著性检验失效。设一元线性回归模型为 $Y_i + \beta_1 + \beta_2 x_i + \mu_i$ 在 $H_0 (\beta = 0)$ 成立下, β_2 的参数估计量 $\hat{\beta}_2$ 的 t 统计量为 $t^* = \hat{\beta}_2 / se(\hat{\beta}_2)$ 。当 μ_i 存在异方差时, OLS 估计式不再具有最小方差, 如果仍用 $var(\hat{\beta}_2) = \sigma^2 \sum x_i^2$ 去估计其方差, 会低估其真实方差, 从而低估 $se(\hat{\beta}_2)$, 也就夸大了所估计参数的统计显著性; ③预测精度降低。由于异方差的存在, 使得参数估计值的方差低估了其真实方差, 造成参数的区间估计失真, 进一步影响 Y 的预测期间, 使其精度降低。因而, 异方差的侦查就显得非常重要。

1 对异方差已有检验方法的评述

对线性模型异方差的常见检验方法有: 图示检验法、等级相关系数法、Glejser 检验, Battlett 检验、Breusch-Pagan 检验、Goldfeld-Quandt 检验、Wald 检验、拉格朗日乘数检验、似然比检验和 White 大样本检验等。这些检验方法除等级相关系数法外都是参数检验方法, 共同缺点是需要知道参数的分布信息, 要求参数服从一定的分布才能做检验。因而, 在不知道参数分布的情形下, 这些参数检验方法就失效了。等级相关系数法虽能克服以上缺点但其又有易受到模型异常值点、强影响点、高杠杆点等特殊样本数据的影响, 应用效果有限。针对以上方法的缺陷, 本文提出了异方差的另一非参

数检验方法: 游程检验。该方法无须知道参数分布的信息, 无须参数服从正态或其它特定分布, 也不要求大样本, 在小样本下, 游程检验仍可使用。

2 游程检验的原理

游程检验是样本的随机性检验, 其用途很广。例如当我们要考察生产中出现次品出现是随机的, 还是成群的, 一个时间序列是平稳的还是非平稳的, 模型的随机干扰项是否是白噪声等都可以通过游程检验来确定。

2.1 游程的定义

游程: 连续出现的具有相同特征的样本点为一个游程。为帮助理解, 举一个掷硬币的例子。向上掷 n 次硬币, 将硬币正面向上的情形记为 1, 将硬币背面向上的情形记为 0, 则就可得到有关掷硬币结果的由 0 和 1 表示的一串序列如 100011110001110000, 将序列中排在一起的几个 1 或 0 叫做一个游程。如上例, 游程为 1, 000, 1111, 000, 111, 0000。共有 6 个游程, 游程数为 6。

2.2 游程检验的基本原理

我们观测序列的游程数, 发现如果数列的游程数太少, 则该数列存在成群的趋向, 0 或 1 的出现总是成群出现的。例如, 数列 0000000111111111 的游程数为 2, 游程数小, 其数列中 0 和 1 都是成群出现。如果数列的游程数过多, 则该数列具有混合趋向, 0 和 1 总是交替出现。例如数列 01010101010101, 游程数为 14, 其 0 和 1 是交替出

^{*} 收稿日期: 2004-03-22

作者简介: 兰嘉庆 (1981 年生), 男, 研究生; E-mail: yuwanling007@eyou.com

现的，具有混合趋向。因此，只有游程数不小也不大，处于一个合适范围时，该序列才具有随机性。用 R 来表示游程数。则游程 R 的分布为：

$$E(R) = 2 \frac{nm}{m+n} + 1$$
$$\text{var}(R) = \frac{2mn(2mn - n - m)}{(m+n)^2(n+m-1)} + 1$$
$$Z = \frac{R - E(R)}{\sqrt{\text{var}(R)}} \sim N(0, 1)$$

其中 R 为游程数， m 为序列中 0 的总数， n 为序列中 1 的总数。

我们将 H_0 令为序列具有随机性；如序列具有混合的倾向， P 值为右尾概率；如序列具有聚类的倾向， P 值为左尾概率；如序列是非随机的， P 值为较小的左尾概率的 2 倍。

3 游程检验对异方差的检验

假设计量经济模型为

$$Y_i + \beta_1 + \beta_2 x_i + \mu_i$$

步骤 1 对 Y 和 X 的数据做回归拟合并求出残差 μ_i ，则可得残差 μ_i 的按下标 i 取值排列的一个数列。

步骤 2 将 μ_i 取绝对值 $|\mu_i|$ ，同时将 $|\mu_i|$ 按递升或递降次序排列，求出该排列的中位数 μ^* 。

步骤 3 将 μ_i 按按下标 i 取值排列的数列中各个 μ_i 取绝对值，但保持原序列的顺序不变，并用各个位置的 μ_i 的绝对值 $|\mu_i|$ 去减 μ^* （中位数），如果得正值，即 $|\mu_i| > \mu^*$ ，记为 1，如果得负值，即 $|\mu_i| < \mu^*$ ，记为 0。就可得一组由 0 和 1 组成得序列 001101010100101。该序列中 0 表示在该位置上的 μ_i 的绝对值 $|\mu_i|$ 小于 $|\mu_i|$ 的中位数 μ^* ，1 则表示在该位置上的 μ_i 的绝对值 $|\mu_i|$ 大于 $|\mu_i|$ 的中位数 μ^* 。

步骤 4 对序列 001101010100101 进行游程检验。观测序列 001101010100101 的游程数 R ，并求

$$Z = \frac{R - E(R)}{\sqrt{\text{var}(R)}}$$
$$Z = \frac{R - E(R)}{\sqrt{\text{var}(R)}} \sim N(0, 1)$$

进行检验。

对游程检验结果的解释：如果 Z 值过大，落入右尾拒绝区域，则拒绝原假设，接受备择假设，序列是混合的。即序列为 0101010101010101， $|\mu_i|$ 对其 中位数 μ^* 的差忽正忽负，交替出现。也就是说， $|\mu_i|$ 忽大忽小，交替出现。随机误差项 μ_i 波动大小交替出现，具有一定的波动规律性，并不同方差。

如果 Z 值过小，也拒绝原假设，序列具有成群趋向，即 0 和 1 是成群出现的，0000011111100000。也就是说 $|\mu_i|$ 在某一段总是大于其中位数 μ^* ，在另一段时间又总是小于其中位数 μ^* ，随机误差项 μ_i 波动总在一段时间波动内大，一段时间内波动小，具有一定的波动规律性，并不同方差。

现以一例具体说明游程检验地应用。

例 1 用北京市 1978—1998 年人均储蓄与人均收入做游程检验，其中， X 为人均收入（元）， Y 为人均储蓄（元），经分析人均储蓄受人均收入的影响，可建立一元回归模型进行分析。通过 Eviews 拟和线形模型，则线形模型为

$$Y = -2185.998 + 1.684158X + \mu$$

现对该模型进行游程检验，以判断其是否具有残差异方差。

步骤 1 根据回归模型求出 y_i ，进而求出残差 μ_i （ $\mu_i = y - y_i$ ）并取绝对值，表示如下：

年份	1978	1979	1980	1981	1982
μ_i	1299.00	1189.13	981.67	900.45	749.58
年份	1985	1986	1987	1988	1989
μ_i	628.66	300.13	428.56	126.10	334.03
年份	1983	1984	1990	1991	1992
μ_i	476.78	990.68	978.34	799.23	767.75
年份	1993	1994	1995	1996	1997
μ_i	714.98	1138.74	1781.37	1342.84	887.06
年份	1998				
μ_i	2086.61				

步骤 2 将 μ_i 按升序排列，得中位数 887.06。

步骤 3 用各个年份位置的 μ_i 的绝对值 $|\mu_i|$ 去减 μ^* （中位数），如果得正值，即 $\mu_i > \mu^*$ ，记为 1，如果得负值，即 $\mu_i < \mu^*$ ，记为 0。就可得一组由 0 和 1 组成得序列，序列如下：

年份	1978	1979	1980	1981	1982	1983	1984
	1	1	1	0	0	0	0
年份	1985	1986	1987	1988	1989	1990	1991
	0	0	0	1	1	0	0
年份	1992	1993	1994	1995	1996	1997	1998
	0	1	1	1	1	—	1

步骤 4 序列 11110000000110001111 的游程数是 $R = 5$ 代入

$$Z = \frac{R - E(R)}{\sqrt{\text{var}(R)}}$$

$Z = -2.682$ 故拒绝原假设。该拟和方程

$Y = -2185.998 + 1.684158X + \mu$
存在异方差。

运用 SPSS 所作的游程检验结果如下
Runs Test

	VAR00001
Test Value	887.0607
Cases < Test Value	10
Cases > = Test Value	11
Total Cases	21
Number of Runs	5
Z	-2.682
Asymp. Sig. (2-tailed)	.007

对该题运用 Goldfeld-Quandt 以对比游程检验。
进行 Goldfeld-Quandt 检验得 $\sum e_i^2 = 1372.202$,

$\sum e_i^2 = 5811189; F = \sum e_i^2 / \sum e_j^2 = 4334.937 >$
 $F_{0.05}(6, 6) = 4.28$, 表明随机误差显著地存在异方差。验证了游程检验的结论。

参考文献:

[1] 庞皓, 李南成. 计量经济学[M] . 成都: 西南财经大学出版社, 2001: 95—98.
[2] 古扎拉蒂. 计量经济学[M] . 林少宫, 译. 北京: 中国人民大学出版社, 1999: 358—370.
[3] GREENE W H. Econometric Analysis[M] . 北京: 清华大学出版社, 2001: 507—512.
[4] 高辉. 异方差性的诊断及数据属性影响的实证分析[J] . 山西师范大学学报(自然科学版), 2003, 17(4).
[5] 吴喜之. 非参数统计[M] . 北京: 中国统计出版社, 1999: 47—49.

Runs Test Used for Heteroscedasticity

LAN Jia qing, YU Wan ting

(Graduate School of University of Finance and Economics, Chengdu 610074, China)

Abstract: The present normal heteroscedasticity testing means are parameter means, and they call for a certain distribution of the parameter. However, in practice many parameters do not satisfy those hypothesized distribution. In order to handle the defect of normal heteroscedasticity testing means, we pose a new mean, that is runs test. This method can overcome those defects, and can be also applied to small collection of sample. We introduce the principal of Runs test firstly, then account the use of runs test on heteroscedasticity testing.

Key words: heteroscedasticity; runs test; runs