

多元线性模型异方差检验方法研究

廖桂丽

福建师范大学

DOI:10.32629/ej.v9i5.3498

[摘要] 多元线性回归模型在实际应用中经常会遇到异方差问题。异方差的存在会使回归得到的参数估计量失去有效性并且不能正确进行假设检验,从而导致错误结论。所以有必要对各种异方差检验方法有所了解并能够根据不同情况选用适当的方法是非常重要的。本文主要介绍了三种常用的异方差检验方法: White检验、Breusch-Pagan检验以及Goldfeld-Quandt检验等及其适用范围。

[关键词] 多元线性模型; 异方差性; 回归分析; 检验方法; 模型调整

中图分类号: F760.6 **文献标识码:** A

Research on the Test Methods for Heteroscedasticity in Multivariate Linear Models

Guili Liao

Fujian Normal University

[Abstract] In practical applications, multiple linear regression models frequently encounter heteroscedasticity issues. The presence of heteroscedasticity renders the parameter estimators obtained through regression ineffective and prevents proper hypothesis testing, leading to erroneous conclusions. Therefore, it is crucial to understand various heteroscedasticity testing methods and select the appropriate one based on specific circumstances. This paper primarily introduces three commonly used heteroscedasticity tests—the White test, the Breusch-Pagan test, and the Goldfeld-Quandt test—along with their respective applications.

[Key words] Multivariate Linear Model; Heteroscedasticity; Regression Analysis; Testing Methods; Model Adjustment

引言

在回归分析中,数据方差是否相等对模型拟合优度有重大影响,在多元线性回归模型中,参数估计所基于的前提条件就是误差项方差相等。但是实际上很多情况下数据之间存在着异方差性,这会导致参数估计效率降低并且还会使假设检验失效,所以寻找合适的异方差检验方法以及及时发现这一问题是在回归分析面临的一大难题。比较各种不同的检验方法可以给我们在实际工作中如何选取、改进模型起到一定的指导作用,有利于提高分析结果可靠性和稳健性。

1 异方差性问题的定义与表现

1.1 异方差性的概念

异方差是指回归模型中误差项方差在不同观测值上不一致情况,在多元线性回归模型中,一个基本假定就是误差项方差应相同即同方差。但如果在实际数据中发现误差方差随一些自变量变动而变化,则为异方差。异方差一般出现在样本容量较大或变量之间具有某种非线性关系时,它会降低回归分析可信度与有效性。因此要使回归模型有效,就需要对是否存在以及如何处理异方差进行判断并采取措施。

1.2 异方差性的表现形式

异方差可通过误差项散点图大致判断。即在残差图上,如果随着自变量或者拟合值增大,误差波动越来越大或者越来越小,则说明有异方差;也可以通过对分组数据所做回归模型残差分布来进行观察。如果数据中存在较为明显的非对称或者异方差情况,则很可能就是该原因造成的^[1]。而一些行业的数据,比如金融以及宏观经济等,由于受到外界影响较大或者本身的性质决定也常常会出现异方差的问题,在这种情况下就需要分析师使用相应的检验手段发现并且解决它。

1.3 异方差性对回归模型的影响

异方差性在回归模型中的影响主要体现在参数估计的效率上。存在异方差性时,最小二乘估计(OLS)仍然是无偏的,但不再是最优的,估计量的方差将被低估,从而导致假设检验的结果不可靠。具体来说,标准误差的计算不准确,会导致t检验和F检验的结果偏差,进而影响到回归系数的显著性判断。另一方面,异方差性还可能使得模型的预测能力受限,增加模型的误差,影响最终决策的准确性。在回归分析过程中,发现并解决异方差性问题对确保结论的稳健性至关重要。

2 常见的异方差检验方法

2.1 White检验方法

White检验是用于检验异方差的一种常用的方法, 比较灵活并且适用范围广。它是基于残差平方与自变量之间的关系来判断是否存在异方差的情况。White检验不需要假定误差项遵循某种特定的概率分布, 所以对模型的要求较低, 可以应用于各种不同类型的样本。当样本有异方差时, White检验能很好地发现误差项方差的变化情况。而在进行White检验时一般要计算异方差检验统计量进行判断, 并根据这个统计量是否显著来判断是否存在异方差的问题。这种方法的一个优点是可以考虑自变量之间可能存在相互影响或者非线性关系, 但是也有可能因为过于宽松的假设而导致较大的误差。

2.2 Breusch-Pagan检验方法

Breusch-Pagan检验方法是常用的异方差检验方法, 用于检测回归模型中的误差项方差是否与某个或某些自变量有关, 即通过计算残差平方对自变量进行回归得到的结果来判断是否存在异方差的问题, 在实际操作中由于Breusch-Pagan检验基于线性模型假设而不能很好地处理非线性回归问题, 所以在使用时需适当修正^[2]。虽然这种方法较为简便, 但是其缺点是在自变量与误差项之间有某种形式的非线性联系的情况下, 检验效果较差。Breusch-Pagan检验只适合自变量与误差项之间有线性关系的情形。

2.3 Goldfeld-Quandt检验方法

Goldfeld-Quandt检验是利用数据进行分组的方法来检验异方差问题的一种方法, 在数据中选取一部分作为第一组, 另一部分作为第二组, 然后在这两组中各自进行回归估计, 再对比这两部分回归结果中的残差平方和, 如果这两部分残差方差有较大差异, 则说明存在异方差性。Goldfeld-Quandt检验的优点在于简单易懂, 通过分组对比可以很直观地看出是否存在异方差性的问题。这种方法适合大数据量并且有明显分界点的数据集。但Goldfeld-Quandt检验的前提条件是能将数据合理地分成两部分, 并且对于分组顺序非常敏感, 容易受样本影响。所以在应用时要结合具体情况选取合适的分组方式以保证检验的有效性。

3 异方差检验方法的适用性分析

3.1 不同数据类型的适用性

不同的异方差检验方法适用于不同类型的数据。对于大量经济数据而言, 由于White检验具有高度灵活性以及对模型假定宽松的要求, 在很大程度上可以正确识别出数据中是否存在异方差; 而对于少量数据或者自变量间存在线性关系的数据集, 则Breusch-Pagan检验更为合适, 这是因为该检验是基于残差与自变量之间的相关性来进行异方差检验; 而Goldfeld-Quandt检验只适合那些有明显分界点并且可以被合理划分成两部分的数据集, 在这种情况下, 如果数据量很大而且自变量之间关系明确, 则可以用Goldfeld-Quandt给出简单有效结论; 而对于非线性模型或者有较复杂关系的数据集, White检验更佳, 它可以解决比较复杂的问题以及误差项。

3.2 方法之间的比较

在选取异方差检验方法时, 需要对各种方法的适用性和优缺点进行比较分析。White检验的优点是非参数的方法, 不需要关于误差项的任何假定, 可用于不同类型的数据, 在变量之间有很强非线性时具有很好的稳健性。但是也有一个很大的问题就是可能会有很高的I型错误的概率, 特别是在样本量较小的时候^[3]。而Breusch-Pagan检验简便易懂, 但是对误差项的要求较高, 即认为误差项与自变量之间呈线性关系, 在现实中很难做到这一点。另外Goldfeld-Quandt检验是用分段的方法来判断是否存在异方差问题, 但是这种方法的应用范围有限, 在数据缺乏明显的分组特征时效果不佳。

3.3 选择检验方法的关键因素

在选取合适的异方差检验方法上, 数据的特点以及检验目的起到决定性的作用。而样本量大小及数据分布情况也会影响检验的效果。一般来说, 在大样本情况下, 使用White检验较为合适, 因为不需要满足严格的条件而且适合处理复杂的数据; 但是在小样本的情况下, 由于Breusch-Pagan检验简便易操作并且适合于线性关系, 因此它更受青睐; 如果可以对数据进行分类, 则Goldfeld-Quandt检验可作为一个简便有效的方法。同时, 检验的功效、方法所要求的前提条件以及最后的应用领域等也会对检验方法的选择产生影响, 在做异方差检验时需要全面地考虑这些问题。

4 异方差性检验在实际数据中的应用

4.1 数据准备与预处理

在进行异方差性检验前, 做好数据准备工作十分重要。首先要保证数据真实可靠, 对数据中出现遗漏值或者异常值要妥善解决, 否则会影响最终检验结果可信度。其次是对数值型变量做标准化或者归一化处理, 特别是当自变量单位相差很大时, 标准化可避免因单位不同造成的偏差。然后就是计算残差, 在数据分析工作中这是一个必要步骤, 回归模型要拟合出残差, 残差分析可用于观察误差项分布以及其变化规律。而在处理数据时也应该注意到变量间可能存在多重共线性问题, 用方差膨胀因子(VIF)来检查是否有多重共线性, 并对其进行相应处理, 有利于增强所采用的方法效果。最后还可以利用一些图形来进行直观判断, 比如残差图和散点图都可以给之后进行异方差性检验提供一定参考依据, 帮助发现是否存在异方差性问题。

4.2 检验方法的实施

进行异方差性检验时要根据所选数据类型以及模型特点选用合适的检验方法, 再对数据进行回归之后可根据数据量大小和变量间关系选择相应的检验方法。当样本比较大时, 由于White检验不需要假定数据服从某种分布因此被广泛采用, 适用于非线性回归模型^[4]。而在样本比较小的时候, Breusch-Pagan检验的效果较好, 尤其是当误差项与自变量之间有线性关系时。在操作中可以通过计算检验值然后对其进行假设检验来得出是否有显著性差异。如果检验发现存在异方差问题, 则需要对模型做出相应修正或者变换。但是需要注意的是每种方法都有各自

的优缺点并且在进行检验之前必须满足一定的条件,在此基础上可以选取适当的显著性水平(例如0.05或者0.01)并结合实际情况分析其结果。

4.3 案例分析与结果解读

在实践中,异方差性检验方法通过对具体案例的数据进行分析来发现其中的问题,在案例分析中一般是对不同的数据集进行回归以及对其残差进行分析,然后利用这些方法判断异方差是否成立,在一个具体的经济问题的研究过程中,通过Breusch-Pagan检验得出某行业回归模型中的残差与生产的数量有关,因此可以认为这个数据有异方差性,据此用加权最小二乘法(WLS)对该模型进行修正,使参数估计更有效、准确,而结果也证明这样做的确使得模型具有更好的预测能力,且检验的结果更有说服力;而在另一个实例中,则是通过White检验发现了数据之间的非线性关系,在此基础上对数据进行变换并调整模型后,使模型更为稳健。

5 模型调整与改进策略

5.1 对异方差性的调整方法

一旦发现异方差性之后,要尽快地作出相应措施来解决问题,在所有的方法中最常用的就是加权最小二乘法(WLS),它通过对每一个观测赋予不同权数的方式来处理不同方差的问题,在WLS中,一般把权数设置为与误差项方差呈倒数的关系,这样就使方差大的观测对于回归的影响降低,而方差小的观测就具有更大的影响,在一些特殊情况之下,如果数据之间存在明显非线性关系,则可对数据进行一定的变换,例如对自变量或者因变量取对数或其他形式变换,使其误差项方差尽可能均匀分布,这样不仅能够缓解异方差带来的问题,而且也能提升模型拟合效果以及预测能力。另外一种比较常用的办法即用稳健的标准误来进行调整,这种方法是通过改变标准误来弥补由于异方差所造成的对系数估计值的影响,从而使假定检验更为合理有效。

5.2 模型重估与假设检验的修正

在修正回归模型以消除异方差的影响之后,需要再评估该模型能否有效描述所给的数据集。而在重新评估时应注意两点:一是要基于新的数据以及已修改过的模型来估计回归参数以及其他相关统计量;二是对假设检验也要进行相应的修正工作。因为异方差的存在会使传统t检验和F检验变得无效,在修正后的模型中应用新的方法来进行假设检验,如加权回归后t检验或者使用稳健标准误等^[5]。另外,在重新评估时还需要检查调整后回归系数的有效性即它们在新的模型中是否显著以及模型整体拟合效果是否更好以保证改进后的模型比原来的模型有更好的预

测能力并且有更小的误差。

5.3 实践中的注意事项

在实际数据分析中,在对模型进行修正解决异方差问题中应注意几个方面的问题。一是数据质量对于模型修正的效果起决定作用。缺失值、异常值以及数据偏态都会影响到异方差检验结果及其后续的修正,所以在做模型修正之前要对数据进行清洗和预处理工作做到最好;二是选取合理的检验方法和修正方法很重要,尤其是当面临比较复杂的情况时,要根据具体情况选用最合适的异方差性检验方法。而且不同的回归模型(如线性回归与非线性回归)其对应的修正方式也是不一样的,应该针对具体的情况选用加权最小二乘法、稳健的标准误或者变换数据的方法,在此基础上还要保证模型优化之后具有良好的可解释性和简洁性,以免因为过度修正而导致模型变得过于复杂从而难以理解和利用。

6 结语

异方差问题是多元线性回归分析的重要内容之一,该问题可以提高模型精度和鲁棒性。采取不同检验方法及修正措施可使回归效果大大改进,在实际工作中根据实际情况选用适当检验方法及修正措施十分必要。

[基金项目编号]

福建省自然科学基金:2022J01191。

国家自然科学基金:12301335。

[参考文献]

[1]徐博韬,余鹏,韩博.多元线性回归模型下国有股权参股对家族企业“漂绿”行为的影响[J].重庆理工大学学报(自然科学),2025,39(4):239-246.

[2]杨筱菡,杨国庆.多元线性回归方程的显著性检验[J].大学数学,2025,41(5):84-90.

[3]刘红彩,于振海,张龙岗.多元线性回归分析在水产养殖中的应用及其在SPSS上的实现[J].黑龙江水产,2025,44(4):424-429.

[4]仲雅娟,李兴才.基于多元线性回归模型的光伏发电项目资本金内部收益率测算研究[J].上海节能,2025(1):116-123.

[5]谢雨雨.基于多元线性回归模型的江苏农村居民可支配收入增长分析[J].广东蚕业,2025,59(7):69-71.

作者简介:

廖桂丽(1983--),女,汉族,福建建瓯人,博士,职称:讲师,单位:福建师范大学,研究方向:时间序列分析。