

# Ljung-Box检验及其改进方法的比较研究

孙淇铭

天津商业大学经济学院, 天津

收稿日期: 2023年3月26日; 录用日期: 2023年4月21日; 发布日期: 2023年4月28日

## 摘要

在时间序列模型的建模过程中, 需要对模型拟合的充分性进行检验, 通常会检验是否存在残差自相关。目前, Ljung-Box检验是被广泛应用于各种时间序列模型建模分析中, 检验残差序列是否存在自相关的方法。该方法被证实存在一些缺陷, 后继学者们提出了一些新的改进方法。选取了四种基于Ljung-Box检验改进方法, 通过蒙特卡洛模拟对其有限样本情况下的性质和表现进行了研究, 指出了这些方法在实际应用中需要注意的问题。

## 关键词

时间序列模型, Ljung-Box检验, 蒙特卡洛模拟, 检验功效

# Comparative Study of Ljung-Box Test and Its Improved Methods

Qiming Sun

School of Economics, Tianjin University of Commerce, Tianjin

Received: Mar. 26<sup>th</sup>, 2023; accepted: Apr. 21<sup>st</sup>, 2023; published: Apr. 28<sup>th</sup>, 2023

## Abstract

In the modeling process of time series model, it is necessary to test the adequacy of model fitting, usually to test whether there is residual autocorrelation. At present, Ljung-box test is widely used in modeling and analysis of various time series models to test whether there is autocorrelation in residual series. This method was proved to have some defects, and some new improvement methods were put forward by subsequent scholars. Four improved methods based on Ljung-box test are selected, and their properties and performance in the case of finite samples are studied by Monte Carlo simulation, and the problems needing attention in practical application of these methods are pointed out.

## Keywords

Time Series Model, Ljung-Box Test, Monte Carlo Simulation, Test Efficacy

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在时间序列模型的建模分析过程中，诊断检验是非常重要的步骤，具体体现在检验拟合模型的残差序列是否存在自相关。

假设时间序列  $\{Y_t\}$  由平稳可逆的自回归移动平均模型  $ARMA(p,q)$  生成，具体公式为：

$$\phi(L)y_t = \theta(L)\varepsilon_t \quad (1)$$

其中  $\phi(L) = 1 + \phi_1 L + \dots + \phi_p L^p$ 、 $\theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q$ ， $L$  为滞后算子， $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ 。则  $\hat{\alpha}, \hat{\theta}, \hat{\varepsilon}$  为估计值，许多诊断检验方法的提出基于残差的自相关系数，

$$r_k = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2} \quad (2)$$

其中  $k = 1, 2, 3, \dots$ 。

最初，Box 和 Pierce (1970) [1] 用残差的自相关系数构造了如下统计量：

$$Q = n \sum_{k=1}^m r_k^2 \quad (3)$$

$m$  是选取的残差自相关系数的个数。在残差序列是白噪声序列 ( $r_k = 0$ ) 的原假设下， $Q$  近似服从自由度为  $m$  的  $\chi^2$  分布。使用这个统计量，明显可以发现较多的残差自相关系数将导致较大的  $Q$  值。Ljung 和 Box (1978) [2] 在研究中提出了更优且对小样本同样适用的修正的  $Q$  统计量，它的结构如下：

$$Q_{LB} = n(n+2) \sum_{k=1}^m (n-k)^{-1} r_k^2 \quad (4)$$

在原假设条件下， $Q_{LB}$  近似服从自由度为  $m-p-q$  的  $\chi^2$  分布。

Ljung-Box 检验(简称为 L-B 检验)自提出以来被广泛应用于实证研究中，但是，一些学者在后续的研究过程中发现 L-B 检验本身存在局限性并在实际检验中表现不佳，例如 McLeod (1978) [3] 证明其统计量的卡方渐近分布依赖于某种特定的线性形式；Lee (2016) [4] 发现对于波动性建模，由于条件方差的存在会使 L-B 检验的实际检验水平发生严重扭曲。对于此类针对模型识别结果的检验过程，如果在检验上出现偏差，就很有可能造成模型识别错误，进而影响后续对实证问题的分析与研究。因此，实证工作者应引起注意。

由于 L-B 检验容易理解且大多数统计软件都提供了该统计量的计算结果，使得有关时间序列分析的实证研究过分依赖于 L-B 检验。但是，随着理论研究的进展，基于 L-B 检验所发展出来的新的检验不断被提出，并且在计算和使用上也比较便利。本文选取的四种基于 L-B 检验的改进方法，包括 Monti 检验、Mahdi-McLeod 检验以及 Fisher 和 Gallagher 提出的两种检验方法。它们基于残差的自相关系数或偏相关系数构建检验统计量。从计算方面来说，这两种相关系数均可以在常用的统计软件中提取，改进方法不需要进行额外的计算；从有限样本的渐近分布来看，新的检验统计量都服从常用分布，也不需要重新计算

检验的临界值。

此外, L-B 检验以及本文选取的四种改进方法中, 都需要自行设定  $m$  的取值。针对这一问题, Ljung (1986) [5] 提出对于常用的时间序列  $m = 5$  较为合适。Battaglia (1990) [6] 研究了  $m$  对 L-B 检验的影响, 提出  $m = N^{1/2}$  更为合适。Hyndman (2014) [7] 从残差自相关系数中包含的信息的角度出发, 认为  $m$  应该足够大, 并提出对于非季节性时间序列  $m = \min(10, N/5)$ , 对于季节性时间序列  $m = \min(2h, N/5)$ ,  $h$  为季节性滞后。Hassani 和 Yeganegi (2019) [8] 从理论上证明了  $m$  的选取会对检验造成影响。此后, Hassani 和 Yeganegi (2020) [9] 进一步在先满足实际检验水平不高于名义检验水平的前提下, 研究使检验功效尽可能大的  $m$  的取值, 得出在样本容量低于 500 时,  $m$  取 3。分析以上  $m$  的取值建议可以发现, 各位学者都考虑了小样本情况的  $m$  的选取问题。但是在其文章和其他的相关研究的模拟实验中, 小样本情况并没有被充分的考虑到。针对  $m$  的取值, 本文考虑到相关研究的常用取值方法, 以及不同样本量应取不同值, 本文选择 Battaglia (1990) [6] 的方法。

综上, 本文着重研究有限样本条件下, 各个检验方法的性能。样本大小的界定按照常规的界定方法,  $N \leq 30$  为小样本,  $N > 30$  为大样本。首先, 使用蒙特卡罗模拟实验法, 从检验功效与实际检验水平的角度对其进行比较分析, 筛选检验效果更佳的检验方法, 为实证工作提供参考; 其次, 从实际检验水平的角度研究小样本情况下各检验方法对  $m$  的敏感性; 最后, 进行总结性阐述。文章安排如下, 第二部分简单介绍了四种改进方法的检验统计量推导过程和对应的分布函数。第三部分利用蒙特卡罗模拟的方法, 在有限样本的情况下, 对五种检验统计量的实际检验水平进行了对比研究; 同时研究了小样本情况下, 不同  $m$  对实际检验水平的影响; 并对模拟的异常情况进行了分析与解释。第四部分模拟了有限样本条件下, 各个检验统计量的检验功效。第五部分进行总结。

## 2. 四种基于 L-B 检验的改进方法的介绍

### 2.1. Monti 检验

Monti (1994) [10] 提出用残差的偏自相关系数代替自相关系数。两种相关系数大同小异, 在计算相关性时偏自相关系数移除了中间变量的间接影响。如果残差项为白噪声过程, 则残差的偏自相关系数应与零没有显著差异。因此模型拟合的充分性检验可与基于统计量:

$$Q_{MT} = n(n+2) \sum_{k=1}^m (n-k)^{-1} \hat{\pi}_k^2 \quad (5)$$

令  $r_{(m)} = (r_1 r_2 \cdots r_m)'$ ,  $\pi_{(m)} = (\pi_1 \pi_2 \cdots \pi_m)'$ 。通过 Durbin-Levinson 算法, 向量  $\pi_{(m)}$  可以用  $r_{(m)}$  的函数表示,  $\pi_{(m)} = \psi(r_{(m)})$ , 对于第  $k$  个元素而言:

$$\pi_k = \psi_k(r_{(k)}) = \frac{r_k - r_{(k-1)}' R_{(k-1)}^{-1} r_{(k-1)}^*}{1 - r_{(k-1)}' R_{(k-1)}^{-1} r_{(k-1)}} \quad (6)$$

其中,  $R_{(k)}$  是  $k \times k$  的托普利茨矩阵,  $R_{(k)} = (r_{|i-j|})_{i,j=1,2,\dots,k}$ ,  $r_{(k)}^* = (r_k r_{k-1} \cdots r_1)'$ 。  $\hat{\pi}_{(m)} = (\hat{\pi}_1 \hat{\pi}_2 \cdots \hat{\pi}_m)'$  是  $m$  个残差偏自相关系数组成的向量, 然后将式(6)中的误差的偏自相关系数替换为残差的偏自相关系数就可以得到  $\hat{\pi}_k$ , 对于残差的滞后  $k$  阶的偏自相关系数就是指剔除了中间  $k-1$  残差项的干扰后,  $\varepsilon_{t-k}$  对  $\varepsilon_t$  影响的相关度量。Monti 证明了  $\hat{\pi}_k$  的近似分布为均值为 0, 方差为  $(n-k)/(n(n+2))$ 。进而  $Q_M$  服从自由度为  $m-p-q$  的卡方分布。

### 2.2. Mahdi-Mcleod 检验

Mahdi 和 Mcleod (2012) [11] 在将 Pena 和 Rodriguez (2002, 2006) [12] [13] 提出的检验方法拓展至多

元时间序列模型的同时，对其在单变量时间序列模型下的检验统计量进行了修正。

Pena 和 Rodriguez (2002) [12]指出估计的残差可以被认为是来自某个分布的多元数据的样本。在多元分析中，检验一组正态随机变量是否具有标量协方差矩阵的似然比检验与多元变量相关矩阵的行列式成正比。对于平稳的时间序列， $m$  阶的残差相关矩阵可以表示为：

$$\hat{\mathbf{R}}_m = \begin{bmatrix} 1 & r_1 & \cdots & r_m \\ r_1 & 1 & \cdots & r_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_m & r_{m-1} & \cdots & 1 \end{bmatrix} \quad (7)$$

残差自相关的检验统计量可以由  $|\hat{\mathbf{R}}_m|$  的一种转换形式构造，这样可以获得在零假设下更简单的分布形式。检验统计量结构如下：

$$Q_{PR} = n \left( 1 - |\hat{\mathbf{R}}_m|^{1/m} \right) \quad (8)$$

该统计量近似服从伽马分布，为了使该统计量的分布更加近似于伽马分布，将矩阵中的每个自相关系数进行了标准化， $\tilde{r}_k^2 = \frac{n+2}{n-k} r_k^2$ 。得到新的检验统计量：

$$\tilde{Q}_{PR} = n \left( 1 - |\tilde{\mathbf{R}}_m|^{1/m} \right) \quad (9)$$

后续的研究中 Lin 和 Mcleod (2006) [14]强调了自相关系数矩阵  $\hat{\mathbf{R}}_m$  是非负定的，但将其中的自相关系数标准化后，新的矩阵  $\tilde{\mathbf{R}}_m$  不一定是非负定的。而且根据 Pena 和 Rodriguez (2002) [12]得到的统计数据经常是无定义的，他们建议使用蒙特卡洛模拟构造分布的临界值。虽然，这将会使检验过程更加复杂。

该统计量的一种理解可以是式(7)的递归表达式，如果定义  $\hat{r}_{(m)} = (r_1 r_2 \cdots r_m)'$ ，则  $|\hat{\mathbf{R}}_m| = \begin{bmatrix} 1 & \hat{r}'_{(m)} \\ \hat{r}_{(m)} & \hat{\mathbf{R}}_{m-1} \end{bmatrix}$ 。

利用分块矩阵行列式的性质： $|\hat{\mathbf{R}}_m| = |\hat{\mathbf{R}}_{m-1}|(1 - R_m^2)$ ，其中  $R_m^2 = \hat{r}'_{(m)} \hat{\mathbf{R}}_{m-1}^{-1} \hat{r}_{(m)}$  是残差的多重相关系数的平方。通过多次重复使用这个性质，可以得到：

$$|\hat{\mathbf{R}}_m|^{1/m} = \left[ \prod_{i=1}^m (1 - R_i^2) \right]^{1/m} \quad (10)$$

$1 - R_i^2$  是依赖性的一种度量， $|\hat{\mathbf{R}}_m|^{1/m}$  是这些度量的几何平均值。因此  $1 - |\hat{\mathbf{R}}_m|^{1/m}$  可以解释为平均的平方相关系数。再根据方差分析中残差平方和的性质可以得到，

$$\frac{1 - R_1^2}{1 - R_{i-1}^2} = \frac{RSS(1, i)}{RSS(1, i-1)} = 1 - \hat{\tau}_i^2 \quad (11)$$

式(10)和式(11)联立可解得：

$$|\hat{\mathbf{R}}_m|^{1/m} = \prod_{i=1}^m (1 - \hat{\tau}_i^2)^{(m+1-i)/m} \quad (12)$$

Mahdi 和 Mcleod (2012) [11]提出了自相关系数的托普利茨矩阵的行列式的对数形式，利用式(12)的结果，提出检验统计量：

$$Q_{mm} = \frac{-3n}{2m+1} \log |\hat{\mathbf{R}}_m| = -3n \sum_{i=1}^m \frac{m+1-i}{m+1} \log(1 - \hat{\tau}_i^2) \quad (13)$$

使用近似于 Pena 和 Rodriguez (2002, 2006) [12] [13]中的方法以及 Box (1954) [15]的一个重要结果，

他们证明了  $Q_{mm}$  的渐近分布近似为自由为  $(3/2)m(m+1)/2m+1-(p+q)$  的卡方分布。

### 2.3. Fisher-Gallagher 检验

Fisher 和 Gallagher (2012) [16] 利用高维数据分析的思想, 推导出新的基于  $m$  阶自相关矩阵的迹的平方的组合检验。

假设  $\Sigma_m$  表示矩阵  $|\hat{R}_m|$  的概率极限, 为了检验独立同分布的相关结构, 检验  $H_0: \Sigma_m = I$ , 这等价于检验  $\Sigma_m$  中的每一个特征值是否为 1。在实践中, 可以使用矩阵  $|\hat{R}_m|$  和它的特征值  $\alpha_i (i = 1, \dots, m+1)$  检验  $H_0$ 。似然比准则使用特征值的几何平均值。如果矩阵  $|\hat{R}_m|$  是奇异或接近奇异的, 则可以提出基于特征值的算术平均值的统计量, 其中需注意  $\frac{1}{m+1} \sum_{k=1}^{m+1} (\alpha_k - 1)^2 \geq 0$ , 通过一些代数转化, 可以等价于:  $\frac{1}{m+1} \text{tr}(\hat{R}_m^2) - 1 \geq 0$ 。再通过计算矩阵  $\hat{R}_m^2$  的迹, 可以考虑基于不等式  $\sum_{k=1}^m \frac{2(m-k+1)}{m+1} \hat{r}_k^2 \geq 0$  对假设进行检验。

不等式两边同乘  $(m+1)/2m$ , 再对残差的自相关系数进行标准化, 可以得到统计量:

$$Q_{WLB} = n(n+2) \sum_{k=1}^m \frac{m-k+1}{m} \frac{\hat{r}_k^2}{n-k} \quad (14)$$

该统计量可以理解为一个加权的 Ljung-box 统计量, 滞后 1 期的残差自相关系数的权重为 1, 滞后  $m$  期的权重为  $1/m$ 。使用残差的偏自相关系数进行类似的推导, 可以得到一个加权的 Monit 统计量:

$$Q_{WMT} = n(n+2) \sum_{k=1}^m \frac{m-k+1}{m} \frac{\hat{\pi}_k^2}{n-k} \quad (15)$$

为了便于计算, 使用与 Pena 和 Rodriguez (2002) [12] 相同的理论与方法, 得到统计量的近似分布为伽马分布, shape 和 scale 参数分别为:

$$\alpha = \frac{3}{4} \frac{m(m+1)^2}{2m^2 + 3m + 1 - 6m(p+q)} \quad (16)$$

$$\beta = \frac{2}{3} \frac{2m^2 + 3m + 1 - 6m(p+q)}{m(m+1)} \quad (17)$$

通过介绍可以看出,  $Q_{MT}$  与  $Q_{LB}$  在零假设的条件下, 是渐近等价的, 但因为相关系数的选择不同, 它们在小样本下的表现会不同;  $Q_{MM}$  的渐近分布具有比较保守的自由度, 这会使它不容易犯第一种错误;  $Q_{WLB}$  和  $Q_{WMT}$  是给  $Q_{LB}$  和  $Q_{MT}$  加上了权重, 滞后期较前的相关系数权重较大, 随着滞后期的增加, 同期的残差自相关系数的权重越小。本文主要研究有限样本下的各检验方法的性能, 具体四种改进方法的优劣, 在下文中进行模拟分析。

### 3. 有限样本下的实际检验水平的模拟研究与结果分析

实际检验水平就是犯第一类错误的概率。通过蒙特卡罗模拟的方法, 在原假设的基础上模拟生成数据, 之后进行检验, 重复多次, 拒绝原假设的概率就是实际检验水平。Davidson 和 MacKinnon (1999) [17] 介绍了另一种与实际检验水平有关的方法是水平扭曲, 它指实际检验水平与名义检验水平的差值, 用来刻度极限样本分布能否更好的近似于有限样本分布。

接下来, 本文将对  $Q_{LB}$ ,  $Q_{MT}$ ,  $Q_{MM}$ ,  $Q_{WLB}$  和  $Q_{WMT}$  (分别对应公式(4)、式(5)、式(13)、式(14)、式(15)) 四种检验方法的实际检验水平与检验功效进行研究。在本章的模拟实验中, 重复生成 1000 个时间序列, 其

中样本量考虑了 $N=20、30、50、80、100、500$ 六种情况， $m = N^{1/2}$  (选取近似的整数值)。

实际检验水平的蒙特卡罗模拟实验的具体步骤为：

1) 生成数据：以AR(1)模型

$$y_t = \varphi y_{t-1} + \varepsilon_t$$

作为数据生成过程，生成时间序列数据。其中考虑到系数对模拟结果的影响，借鉴 Pena 和 Rodriguez (2002) [12]中的模型， $\varphi$  取 0.1、0.3、0.5、0.7、0.9；

2) 拟合模型：用生成的数据再次拟合AR(1)模型；

3) 进行检验：用五种不同的方法在给定显著性水平 $\alpha = 0.05$ 的条件下，对残差序列进行自相关检验；

4) 重复实验：将步骤1~3重复1000次，计算拒绝原假设的概率。

### 3.1. 有限样本下的实际检验水平的模拟

表 1 展示了不同的有限样本下，基于 AR(1)模型的实际检验水平的模拟结果。采用 Battaglia (1990) [6] 中  $m$  的取值方法，针对不同的  $N=20、30、50、80、100、500$ ， $m=4、5、7、9、10、22$ 。

在  $\varphi = 0.9$  的模型中， $N=20、30$  时，实际检验水平过大； $N=50$  时各检验统计量的实际检验水平都大于 0.05。原因可能为在模型估计的系数越接近 1 时，估计越容易出现异常，进而模拟结果出现异常。本节的结果分析中，先抛开此三种情形，后续单独对异常情况进行分析和解释。

**Table 1.** Actual test level of five test methods based on AR (1) model with limited samples ( $\alpha = 0.05$ )

**表 1.** 基于 AR(1)模型的有限样本下五种检验方法的实际检验水平模拟分析( $\alpha = 0.05$ )

$\varphi$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$
$N = 20 \quad m = 4$						$N = 30 \quad m = 5$				
0.1	0.040	<b>0.069</b>	0.020	0.039	0.040	0.049	<b>0.058</b>	0.027	0.038	0.033
0.3	0.037	<b>0.076</b>	0.026	0.038	0.034	0.040	0.045	0.013	0.039	0.038
0.5	<b>0.055</b>	<b>0.069</b>	0.033	0.035	0.044	<b>0.053</b>	<b>0.059</b>	0.023	0.037	0.035
0.7	0.050	<b>0.059</b>	0.028	0.041	0.048	0.043	0.048	0.029	0.034	0.039
0.9	<b>0.965</b>	<b>0.965</b>	<b>0.967</b>	<b>0.966</b>	<b>0.968</b>	<b>0.297</b>	<b>0.301</b>	<b>0.284</b>	<b>0.293</b>	<b>0.297</b>
$N = 50 \quad m = 7$						$N = 80 \quad m = 9$				
0.1	<b>0.055</b>	<b>0.066</b>	0.036	0.031	0.042	0.045	0.049	0.028	0.032	0.035
0.3	0.046	<b>0.053</b>	0.023	0.035	0.038	0.043	0.035	0.019	0.031	0.043
0.5	0.043	0.045	0.018	0.047	0.039	<b>0.060</b>	<b>0.057</b>	0.031	0.042	0.036
0.7	<b>0.064</b>	<b>0.063</b>	0.034	0.044	0.044	<b>0.057</b>	<b>0.054</b>	0.034	0.037	0.043
0.9	<b>0.068</b>	<b>0.082</b>	<b>0.061</b>	<b>0.068</b>	<b>0.067</b>	0.049	0.050	0.040	0.042	0.043
$N = 100 \quad m = 10$						$N = 500 \quad m = 22$				
0.1	0.038	0.037	0.036	0.034	0.038	0.043	0.035	0.042	0.050	0.042
0.3	0.026	0.031	0.021	0.033	0.035	0.039	0.030	0.033	0.047	0.043
0.5	0.040	0.036	0.027	0.035	0.030	0.049	0.047	0.046	0.047	0.044
0.7	<b>0.055</b>	0.035	0.033	0.039	0.036	<b>0.053</b>	0.039	0.039	0.047	0.048
0.9	0.037	0.036	0.039	0.046	0.044	0.040	0.039	0.034	0.049	0.044

注：实际检验水平超过 0.05 的加粗表示。

从实际检验水平的角度分析。1)  $Q_{MT}$  统计量的实际检验水平在样本量  $N = 20, 30, 50, 80$  时的表现差于  $Q_{LB}$ , 经常超过名义检验水平。随着样本容量的增加, 表现逐渐变好; 2)  $Q_{MM}$  统计量的实际检验水平相对比较保守, 全部低于实际检验水平, 且多数低于其他检验统计量; 3)  $Q_{LB}$  统计量的实际检验水平在不同的样本量的情况下, 都会出现大于名义检验水平的情况。然而,  $Q_{WLB}$  和  $Q_{WMT}$  统计量的实际检验水平全部低于且接近名义检验水平。这表明,  $Q_{WLB}$  和  $Q_{WMT}$  统计量与  $Q_{LB}$  相比, 有限样本分布更好的近似于渐近分布, 且在使用时能更少的犯“弃真”的错误; 4) 在去除掉异常情况的前提下,  $Q_{LB}$  统计量实际检验水平的范围是 0.026~0.064, 对比来看,  $Q_{WLB}$  和  $Q_{WMT}$  统计量的范围分别是 0.031~0.050 和 0.33~0.48, 后两者更加稳定。从水平扭曲的角度分析。表 2 展示了去除掉所有  $\varphi = 0.9$  的模拟结果后, 各样本容量下五种检验统计量的平均水平扭曲程度, 该值是在每一个样本中, 将所有模型的水平扭曲值简单平均得到。结果显示,  $Q_{MM}$  统计量的水平扭曲程度较大, 五种样本容量下的平均水平扭曲程度超过 0.015, 即在有限样本下, 使用  $Q_{MM}$  统计量进行残差自相关检验时, 更有可能产生错误的结论。 $Q_{MT}$  统计量在  $N = 20$  时, 扭曲程度较大。除此之外,  $Q_{LB}$ 、 $Q_{WLB}$  和  $Q_{WMT}$  统计量的平均水平扭曲程度全部小于 0.015, 即从此角度分析, 三种检验统计都比较有效。

**Table 2.** The average level of distortion of the five test statistics for each sample

**表 2.** 各样本的五种检验统计量的平均水平扭曲程度

$N$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$
20	0.0070	<b>0.0183</b>	<b>0.0233</b>	0.0118	0.0085
30	0.0053	0.0040	<b>0.0270</b>	0.0130	0.0138
50	0.0075	0.0093	<b>0.0223</b>	0.0108	0.0093
80	0.0060	0.0054	<b>0.0196</b>	0.0132	0.0100
100	0.0128	0.0150	<b>0.0188</b>	0.0126	0.0134
500	0.0072	0.0120	0.0112	0.0022	0.0058

注: 平均水平扭曲程度超过 0.015 的加粗表示。

### 3.2. 小样本情况下实际检验水平的模拟与结果分析

有关时间序列分析检验方法有效性的相关研究中, 很少有详细考虑到小样本条件下各检验统计量性能的情况。因此, 为了判断改进方法在小样本条件下是否稳定, 选取了  $m = 2, 3, 4, 5$  (基于 Battaglia (1990) [6]、Hyndman (2014) [7] 和 Hassani 和 Yeganegi (2020) [9] 的研究结果) 以此来探究小样本情况下,  $m$  的取值是否对模拟过程造成影响。

表 3 展示了, 小样本条件下 ( $N = 20, 25$ ) 基于 AR(1) 模型的各检验统计量的模拟结果。通过计算实际检验水平随不同  $m$  值的变化率, 发现五种检验方法的实际检验水平随  $m$  的变化没有明显的规律性变动。 $Q_{LB}$  统计量的实际检验水平经常出现超过名义显著性水平和  $Q_{MT}$  统计量的实际检验水平在小样本情况下全部超过名义显著性水平的情况与表 2 一致。

而且, 可以发现当  $\varphi = 0.9$  时, 各检验统计量的实际检验水平出现异常的情况并没有随  $m$  的变化出现改善。因此可以得出异常情况的出现与  $m$  取值无关的结论。进一步分析, 发现随着样本容量的提升, 模拟的实际检验水平有明显的改善, 如图 1 所示。分析其原因为, 蒙特卡罗模拟数据有随机性, 当样本容量较小时, 随机性的特点被放大。容易造成小样本的时间序列数据不具备平稳性, 进而造成实际检验水平的模拟结果出现异常。

**Table 3.** Actual test levels of five test methods for small samples based on AR (1) model ( $\alpha = 0.05$ )  
**表 3.** 基于 AR(1)模型的小样本的五种检验方法的实际检验水平模拟分析结果( $\alpha = 0.05$ )

$\varphi$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$
	$N = 20$					$N = 25$				
	$m = 2$					$m = 2$				
0.1	<b>0.067</b>	<b>0.076</b>	0.028	0.038	0.040	0.050	<b>0.057</b>	0.020	0.021	0.024
0.3	0.049	<b>0.061</b>	0.021	0.031	0.032	<b>0.064</b>	<b>0.069</b>	0.021	0.020	0.024
0.5	<b>0.058</b>	<b>0.080</b>	0.034	0.039	0.040	0.050	<b>0.068</b>	0.028	0.025	0.037
0.7	0.049	<b>0.072</b>	0.036	0.044	0.048	<b>0.065</b>	<b>0.076</b>	0.054	0.049	0.050
0.9	<b>0.939</b>	<b>0.940</b>	<b>0.939</b>	<b>0.940</b>	<b>0.940</b>	<b>0.951</b>	<b>0.951</b>	<b>0.952</b>	<b>0.953</b>	<b>0.952</b>
	$m = 3$					$m = 3$				
0.1	0.043	<b>0.056</b>	0.020	0.021	0.023	<b>0.051</b>	0.055	0.020	0.025	0.029
0.3	<b>0.060</b>	<b>0.084</b>	0.035	0.039	0.043	0.040	0.045	0.014	0.017	0.022
0.5	<b>0.058</b>	<b>0.069</b>	0.029	0.036	0.042	<b>0.058</b>	<b>0.067</b>	0.029	0.034	0.039
0.7	<b>0.056</b>	<b>0.066</b>	0.038	0.044	0.050	0.047	<b>0.058</b>	0.033	0.034	0.047
0.9	<b>0.794</b>	<b>0.797</b>	<b>0.782</b>	<b>0.790</b>	<b>0.794</b>	<b>0.920</b>	<b>0.923</b>	<b>0.921</b>	<b>0.920</b>	<b>0.922</b>
	$m = 4$					$m = 4$				
0.1	0.040	<b>0.069</b>	0.020	0.039	0.040	0.049	<b>0.066</b>	0.030	0.034	0.039
0.3	0.037	<b>0.076</b>	0.026	0.038	0.034	<b>0.052</b>	<b>0.072</b>	0.032	0.034	0.044
0.5	<b>0.055</b>	<b>0.069</b>	0.033	0.035	0.044	0.046	<b>0.058</b>	0.033	0.039	0.043
0.7	0.050	<b>0.059</b>	0.028	0.041	0.048	<b>0.057</b>	<b>0.060</b>	0.035	0.038	0.045
0.9	<b>0.965</b>	<b>0.965</b>	<b>0.967</b>	<b>0.966</b>	<b>0.968</b>	<b>0.995</b>	<b>0.994</b>	<b>0.995</b>	<b>0.995</b>	<b>0.995</b>
	$m = 5$					$m = 5$				
0.1	0.045	0.048	0.028	0.030	0.038	<b>0.053</b>	<b>0.066</b>	0.026	0.026	0.036
0.3	0.044	<b>0.060</b>	0.024	0.033	0.037	0.044	<b>0.070</b>	0.029	0.025	0.042
0.5	<b>0.051</b>	<b>0.058</b>	0.025	0.032	0.037	<b>0.051</b>	<b>0.057</b>	0.023	0.029	0.034
0.7	0.048	<b>0.051</b>	0.024	0.030	0.036	<b>0.068</b>	<b>0.071</b>	0.034	0.046	0.048
0.9	<b>0.954</b>	<b>0.954</b>	<b>0.952</b>	<b>0.953</b>	<b>0.952</b>	<b>0.985</b>	<b>0.985</b>	<b>0.986</b>	<b>0.985</b>	<b>0.986</b>

注：实际检验水平超过 0.05 的加粗表示。

#### 4. 有限样本下的检验功效的模拟研究与结果分析

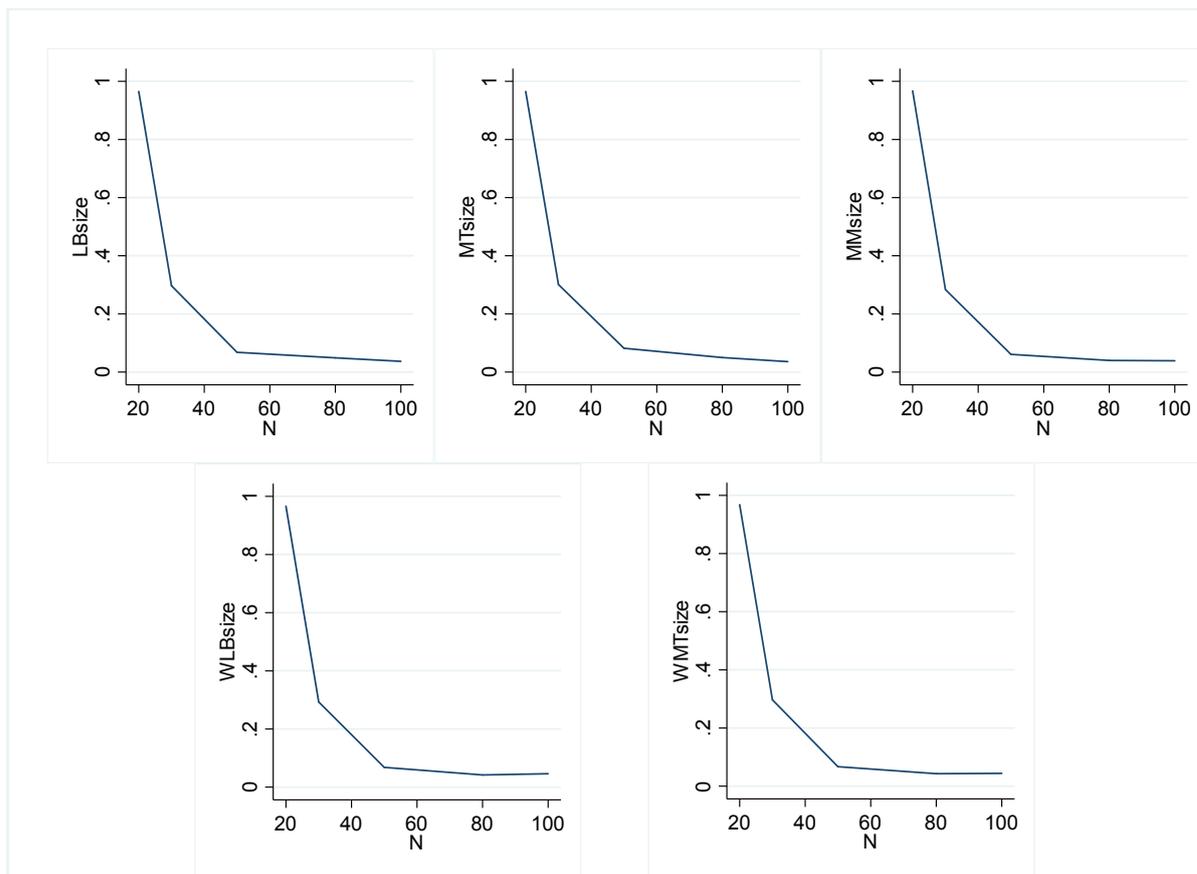
检验功效是指不犯第二类错误的概率，在备择假设的基础上重复生成大量时间序列，然后根据被研究的检验统计量的渐近分布的临界值做出接受或拒绝原假设的判断。这种情况下，拒绝原假设的比率就是该检验统计量的检验功效。

检验功效的蒙特卡罗模拟实验的具体步骤为(以 AR 模型为例)：

1) 生成数据：以 ARMA(2,2)模型

$$y_i = \phi_1 y_{i-1} + \phi_2 y_{i-2} + \varepsilon_i + \theta_1 \varepsilon_{i-1} + \theta_2 \varepsilon_{i-2}$$

作为数据生成过程，生成时间序列数据；



**Figure 1.** The actual test level of five test statistics varies with sample size ( $\varphi = 0.9$ )  
**图 1.**  $\varphi = 0.9$  时五种检验统计量的实际检验水平随样本容量变化图

- 2) 拟合模型：用生成的数据再次拟合 AR(1)和 MA(1)模型；
- 3) 进行检验：用五种不同的方法，在显著性水平  $\alpha = 0.05$  的条件下，对残差序列进行自相关检验；
- 4) 重复实验：将步骤 1~3 重复 1000 次，计算拒绝原假设的概率。

有关检验功效模拟的模型参照了 Monti (1994) [10] 中的设计，考虑了 24 个不同 ARMA(2,2)模型作为数据生成过程。现有的相关研究中样本容量的选择多为 100 和 500，为探究各种有限样本下各检验统计量检验功效的关系，选取  $N = 20, 30, 50, 80, 100$  五种情况， $m$  的取值同样借鉴 Battaglia (1990) [6] 中方法，显著性水平  $\alpha = 0.05$ 。

表 4 展示了在  $N = 20, 30$  的小样本下，各检验统计量的检验功效。结果显示： $Q_{MT}$  检验统计量的检验功效表现良好，全部大于  $Q_{LB}$  统计量的检验功效，且在多数模型中检验功效最大。 $Q_{MM}$ 、 $Q_{WLB}$  和  $Q_{WMT}$  统计量在一些模型中，检验功效低于  $Q_{LB}$  统计量(如模型 1、模型 6~8 等)。计算在各个样本容量下， $Q_{MT}$ 、 $Q_{MM}$ 、 $Q_{WLB}$  和  $Q_{WMT}$  统计量与  $Q_{LB}$  统计量相比全部模型的检验功效平均变化率<sup>1</sup> 分别为 40.4%、-14%、-12.8% 和 10.4% ( $N = 20$ ) 以及 17.5%、-8.3%、-3.3% 和 14.9% ( $N = 50$ )。可见在小样本下， $Q_{MT}$  和  $Q_{WMT}$  统计量的检验效果优于  $Q_{LB}$  统计量。除此之外，在模型 4、模型 6~8 和模型 18~19 中，各统计量的检验功效过小，出现小于 0.05 的情况。分析原因是蒙特卡罗模拟的随机性在小样本情况下被放大；ARMA 模型的可逆性使特定模型生成的数程具备了 AR 模型或 MA 模型的特征。

<sup>1</sup> 检验功效平均变化率：某个样本容量下，24 个模型中，改进统计量与 L-B 统计量的检验功效的差值占 L-B 统计量检验功效的比率的平均值。

表 5 展示了  $N = 50、80、100$  时, 各检验统计量的检验功效。结果显示: 在大样本的条件下,  $Q_{MT}$  检验统计量的优势消失,  $Q_{WMT}$  检验统计的表现更加良好。而且  $Q_{MM}$  和  $Q_{WLB}$  统计量的检验功效与小样本条件下的相比也有所改善, 变得优于  $Q_{LB}$  统计量。与  $Q_{LB}$  统计量相比,  $Q_{MT}$ 、 $Q_{MM}$ 、 $Q_{WLB}$  和  $Q_{WMT}$  统计量的检验功效的平均变化率变为 12.7%、8.6%、2.7%和 15.8% ( $N = 50$ ), 12.6%、16.3%、10.6%和 17.5% ( $N = 80$ ) 以及 10.5%、18.4%、11.5%和 21.0%。可见, 在大样本下, 四种检验统计量均优于  $Q_{LB}$  统计量, 且  $Q_{WMT}$  检验统计最好。模型 4、模型 6~8 和模型 18~19 在小样本条件下的异常情况也随着样本容量的增加有所改善。

Table 4. Simulation of test power for each test statistic based on ARMA (2,2) model I

表 4. 基于 ARMA(2,2)模型的各项检验统计量的检验功效模拟(一)

M	$\varphi_1$	$\varphi_2$	$\theta_1$	$\theta_2$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$
					$N = 20 \quad m = 4$					$N = 30 \quad m = 5$				
模拟数据拟合 AR(1)模型														
1	—	—	-0.50	—	0.118	<b>0.167</b>	0.102	0.108	0.144	0.129	0.144	0.121	0.118	<b>0.148</b>
2	—	—	-0.80	—	0.228	<b>0.381</b>	0.268	0.222	0.333	0.281	0.467	0.433	0.326	<b>0.483</b>
3	—	—	-0.60	0.30	0.116	<b>0.127</b>	0.089	0.104	0.124	0.110	<b>0.113</b>	0.088	0.105	0.110
4	0.10	0.30	—	—	0.045	<b>0.069</b>	0.027	0.037	0.042	0.091	<b>0.117</b>	0.091	0.101	0.113
5	1.30	-0.35	—	—	0.996	0.997	<b>0.998</b>	<b>0.998</b>	<b>0.998</b>	0.977	0.974	0.978	<b>0.979</b>	<b>0.979</b>
6	0.70	—	-0.40	—	0.029	<b>0.041</b>	0.007	0.015	0.013	0.036	<b>0.046</b>	0.026	0.027	0.033
7	0.70	—	-0.90	—	0.067	<b>0.108</b>	0.044	0.044	0.068	0.086	<b>0.116</b>	0.058	0.059	0.074
8	0.40	—	-0.60	0.30	0.049	<b>0.058</b>	0.029	0.031	0.039	0.040	<b>0.052</b>	0.028	0.033	0.036
9	0.70	—	0.70	-0.15	0.843	0.892	0.890	0.856	<b>0.905</b>	0.533	0.745	0.777	0.625	<b>0.812</b>
10	0.70	0.20	0.50	—	0.990	<b>0.993</b>	0.990	0.991	0.991	0.997	0.997	0.997	0.997	0.997
11	0.70	0.20	-0.50	—	0.047	<b>0.071</b>	0.029	0.032	0.044	0.116	0.116	0.107	0.113	<b>0.126</b>
12	0.90	-0.40	1.20	-0.30	0.720	0.830	0.814	0.771	<b>0.858</b>	0.809	0.914	0.938	0.882	<b>0.950</b>
模拟数据拟合 MA(1)模型														
13	0.50	—	—	—	0.044	<b>0.081</b>	0.034	0.036	0.046	0.072	<b>0.089</b>	0.065	0.077	0.087
14	0.80	—	—	—	0.248	0.258	0.232	<b>0.290</b>	0.279	0.459	0.434	0.479	<b>0.530</b>	0.512
15	1.10	-0.35	—	—	0.409	0.436	0.433	<b>0.495</b>	0.487	0.654	0.641	0.676	0.695	<b>0.709</b>
16	—	—	0.80	-0.50	0.078	<b>0.183</b>	0.108	0.064	0.157	0.128	0.233	0.211	0.130	<b>0.244</b>
17	—	—	-0.60	0.30	0.086	<b>0.110</b>	0.071	0.100	0.096	0.115	<b>0.132</b>	0.108	0.119	0.130
18	0.50	—	-0.70	—	0.038	<b>0.064</b>	0.019	0.018	0.031	0.051	<b>0.062</b>	0.031	0.035	0.039
19	-0.50	—	0.70	—	0.029	<b>0.071</b>	0.022	0.018	0.033	0.048	<b>0.060</b>	0.028	0.024	0.036
20	0.30	—	0.80	-0.50	0.051	<b>0.093</b>	0.046	0.027	0.072	0.073	<b>0.100</b>	0.063	0.055	0.086
21	0.80	—	-0.50	0.30	0.185	<b>0.221</b>	0.160	0.192	0.204	0.436	0.478	0.461	0.484	<b>0.503</b>
22	1.20	-0.50	0.90	—	0.901	0.904	0.932	0.943	<b>0.945</b>	0.983	0.981	0.987	0.990	<b>0.989</b>
23	0.30	-0.20	-0.70	—	0.112	<b>0.146</b>	0.096	0.079	0.134	0.137	0.193	0.194	0.155	<b>0.225</b>
24	0.90	-0.40	1.20	-0.30	0.463	0.466	0.480	0.528	<b>0.533</b>	0.615	0.628	0.667	0.689	<b>0.713</b>

注: 每次模拟中, 检验功效最大的加粗表示。

**Table 5.** Simulation of test power for each test statistic based on ARMA (2,2) model II  
**表 5.** 基于 ARMA(2,2)模型的各项检验统计的检验功效模拟(二)

M	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$	$Q_{LB}$	$Q_{MT}$	$Q_{MM}$	$Q_{WLB}$	$Q_{WMT}$
N = 50 m = 7					N = 80 m = 9					N = 100 m = 10					
模拟数据拟合 AR(1)模型															
1	0.177	0.207	0.202	0.180	<b>0.224</b>	0.221	0.252	0.299	0.260	<b>0.311</b>	0.271	0.315	0.391	0.330	<b>0.406</b>
2	0.419	0.712	0.734	0.534	<b>0.758</b>	0.613	0.910	0.957	0.787	<b>0.960</b>	0.739	0.970	0.989	0.899	<b>0.989</b>
3	<b>0.145</b>	0.135	0.104	0.136	0.133	0.178	0.156	0.156	0.189	<b>0.171</b>	0.197	0.181	0.206	0.223	<b>0.216</b>
4	0.159	0.159	0.176	0.183	<b>0.193</b>	0.345	0.315	0.393	0.404	<b>0.398</b>	0.454	0.422	0.532	0.553	<b>0.538</b>
5	0.960	0.961	<b>0.976</b>	0.973	<b>0.976</b>	0.695	0.676	0.813	0.805	<b>0.816</b>	0.847	0.838	0.916	0.914	<b>0.914</b>
6	0.060	0.066	0.041	0.041	<b>0.086</b>	0.080	0.078	0.060	0.070	<b>0.085</b>	0.096	0.095	0.106	0.099	<b>0.110</b>
7	0.102	0.135	0.097	0.069	<b>0.189</b>	0.125	0.188	0.147	0.114	<b>0.195</b>	0.131	<b>0.267</b>	0.234	0.150	0.244
8	0.112	0.136	0.099	0.113	<b>0.174</b>	0.201	0.217	0.252	0.229	<b>0.258</b>	0.221	0.204	0.251	0.252	<b>0.258</b>
9	0.590	0.913	0.955	0.764	<b>0.963</b>	0.955	0.939	0.977	0.983	<b>0.978</b>	0.923	0.999	<b>1.000</b>	0.994	<b>1.000</b>
10	0.926	0.932	0.940	0.932	<b>0.942</b>	0.998	0.998	<b>0.999</b>	<b>0.999</b>	<b>0.999</b>	0.746	0.839	0.911	0.852	<b>0.921</b>
11	0.331	0.318	0.350	0.341	<b>0.378</b>	0.821	0.770	0.820	<b>0.860</b>	0.828	0.720	0.963	<b>0.990</b>	0.887	<b>0.990</b>
12	0.960	0.990	0.995	0.987	<b>0.995</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
模拟数据拟合 MA(1)模型															
13	0.138	0.125	0.137	0.151	<b>0.150</b>	0.228	0.187	0.229	<b>0.249</b>	0.240	0.282	0.260	0.320	0.339	<b>0.335</b>
14	0.775	0.736	0.800	0.820	<b>0.808</b>	0.960	0.942	0.976	<b>0.982</b>	0.976	0.983	0.979	0.989	<b>0.990</b>	0.989
15	0.833	0.887	0.926	0.936	<b>0.929</b>	0.982	0.982	0.991	<b>0.992</b>	0.991	0.999	0.996	1.000	1.000	1.000
16	0.222	0.329	0.370	0.280	<b>0.393</b>	0.388	0.495	0.609	0.512	<b>0.622</b>	0.473	0.584	0.695	0.622	<b>0.704</b>
17	0.182	0.218	0.219	0.213	<b>0.245</b>	0.314	0.353	0.399	0.378	<b>0.411</b>	0.372	0.404	0.517	0.471	<b>0.527</b>
18	0.080	<b>0.077</b>	0.056	0.063	0.062	0.068	<b>0.090</b>	0.071	0.061	0.076	0.101	<b>0.115</b>	0.099	0.100	0.109
19	0.084	<b>0.087</b>	0.057	0.055	0.069	0.105	<b>0.110</b>	0.091	0.089	0.099	0.110	<b>0.117</b>	0.105	0.103	0.114
20	0.079	0.100	0.088	0.075	<b>0.107</b>	0.085	0.087	0.096	0.083	<b>0.106</b>	0.078	0.092	0.102	0.085	<b>0.106</b>
21	0.754	0.772	0.811	0.816	<b>0.824</b>	0.956	0.962	0.986	0.982	<b>0.986</b>	0.988	0.987	0.996	0.996	<b>0.997</b>
22	0.999	0.999	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>							
23	0.266	0.306	0.373	0.330	<b>0.396</b>	0.385	0.428	0.574	0.518	<b>0.594</b>	0.486	0.528	0.667	0.628	<b>0.674</b>
24	0.821	0.837	0.909	0.911	<b>0.919</b>	0.955	0.973	0.992	0.989	<b>0.992</b>	0.987	0.992	0.997	0.994	<b>0.997</b>

注：每次模拟中，检验功效最大的加粗表示。

分析  $N = 20、30、50、80、100$  时，各检验统计量检验功效的模拟结果发现，四种改进的检验统计量的检验功效与  $Q_{LB}$  统计量对比的平均变化率随样本容量的增加呈现出一定的规律，如图 2 所示。 $Q_{MT}$  统计量的检验功效总是大于  $Q_{LB}$  统计量，但随样本容量的增加，此优势逐渐减弱； $Q_{MM}$  和  $Q_{WLB}$  统计量的检验功效在小样本时小于  $Q_{LB}$  统计量，但随样本容量的增加，检验功效逐渐变大，并在  $N = 40$  左右时超过  $Q_{LB}$  统计量； $Q_{WMT}$  统计量的检验功效一直大于  $Q_{LB}$  统计量，且随样本容量的增加而持续变大。因此，小样本时用  $Q_{MT}$  统计量做残差自相关检验更好，大样本时用  $Q_{WMT}$  统计量更优。

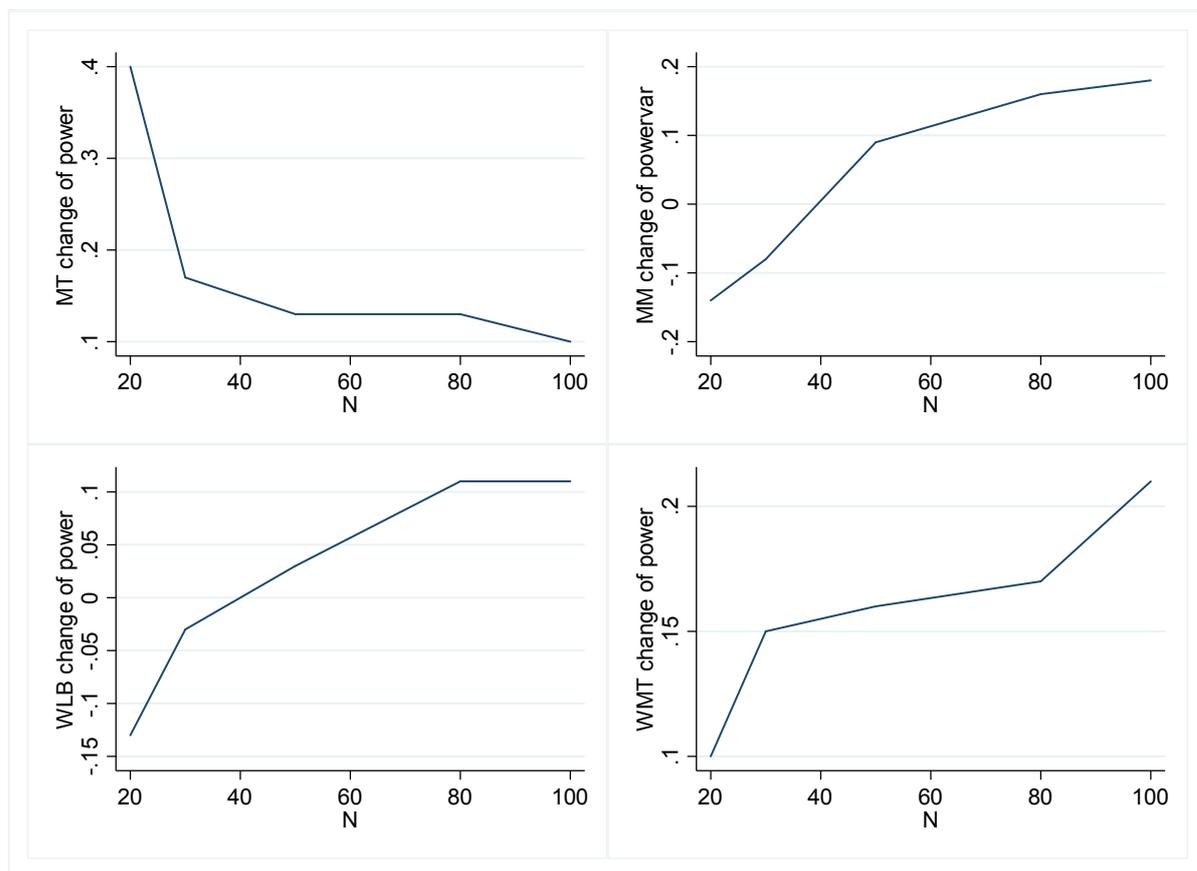


Figure 2. Trend chart of average change rate of four improved statistics compared to L-B statistics

图 2. 四种改进统计量检验功效与 L-B 统计量相比的平均变化率趋势图

## 5. 结论

本文主要进行了 L-B 检验和它的四种改进方法在有限样本条件的实际检验水平与检验功效的模拟分析。

在小样本的条件下( $N \leq 30$ )。只有  $Q_{MT}$  和  $Q_{WMT}$  统计量的检验功效优于  $Q_{LB}$  统计量，但  $Q_{MT}$  统计量的实际检验水平表现较差，经常超过名义检验水平，且水平扭曲程度最严重。

在大样本的条件下( $N > 30$ )。四种改进检验统计量的检验功效均优于  $Q_{LB}$  统计量， $Q_{WMT}$  统计量的表现最好。且  $Q_{MM}$ 、 $Q_{WLB}$  和  $Q_{WMT}$  统计量的实际检验水平均小于名义检验水平，其中  $Q_{WLB}$  和  $Q_{WMT}$  统计量表现更好， $Q_{MM}$  统计量的水平扭曲程度较大。

综上所述，小样本时只有  $Q_{WMT}$  统计量检验效果较好；随着样本容量的增加，其余改进统计量的检验效果逐渐改善。相关理论研究中，小样本情形一直是被忽略的部分。实证研究中，小样本的情况却十分常见。因此，建议实证学者们警惕 L-B 检验在样本较小情形下的应用，可以同时使用  $Q_{WMT}$  统计量进行多次检验，确保模型设定的正确。

## 参考文献

- [1] Pierce, B.D.A. (1970) Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. *Journal of the American Statistical Association*, **72**, 397-402.
- [2] Ljung, G.M. and Box, G.E.P. (1978) On a Measure of Lack of Fit in Time Series Models. *Biometrika*, **65**, 297-303.

- <https://doi.org/10.1093/biomet/65.2.297>
- [3] Mcleod, A.I. (1978) On the Distribution of Residual Autocorrelations in Box-Jenkins Models. *Journal of the Royal Statistical Society*, **40**, 296-302. <https://doi.org/10.1111/j.2517-6161.1978.tb01042.x>
- [4] Lee, T. (2016) Wild Bootstrap Ljung-Box Test for Cross Correlations of Multivariate Time Series. *Economics Letters*, **147**, 59-62. <https://doi.org/10.1016/j.econlet.2016.08.015>
- [5] Ljung, G.M. (1986) Diagnostic Testing of Univariate Time Series Models. *Biometrika*, **73**, 725-730. <https://doi.org/10.1093/biomet/73.3.725>
- [6] Battaglia, F. (1990) Approximate Power of Portmanteau Tests for Time Series. *Statistics & Probability Letters*, **9**, 337-341. [https://doi.org/10.1016/0167-7152\(90\)90143-U](https://doi.org/10.1016/0167-7152(90)90143-U)
- [7] Thoughts on the Ljung-Box Test. <https://robjhyndman.com/hyndsight/ljung-box-test>
- [8] Hassani, H. and Yeganegi, M.R. (2019) Sum of Squared ACF and the Ljung-Box Statistics. *Physica A: Statistical Mechanics and Its Applications*, **520**, 81-86. <https://doi.org/10.1016/j.physa.2018.12.028>
- [9] Hassani, H. and Yeganegi, M.R. (2020) Selecting Optimal Lag Order in Ljung-Box Test. *Physica A: Statistical Mechanics and Its Applications*, **541**, 123700. <https://doi.org/10.1016/j.physa.2019.123700>
- [10] Monti, A.C. (1994) A Proposal for a Residual Autocorrelation Test in Linear Models. *Biometrika*, **81**, 776-780. <https://doi.org/10.1093/biomet/81.4.776>
- [11] Mahdi, E. and Mcleod, A.I. (2012) Improved Multivariate Portmanteau Test. *Journal of Time Series Analysis*, **33**, 211-222. <https://doi.org/10.1111/j.1467-9892.2011.00752.x>
- [12] Pena, D. and Rodriguez, J. (2002) A Powerful Portmanteau Test of Lack of Fit for Time Series. *Journal of the American Statistical Association*, **97**, 601-611. <https://doi.org/10.1198/016214502760047122>
- [13] Peña, D. and Rodríguez, J. (2006) The Log of the Determinant of the Autocorrelation Matrix for Testing Goodness of Fit in Time Series. *Journal of Statistical Planning and Inference*, **136**, 2706-2718. <https://doi.org/10.1016/j.jspi.2004.10.026>
- [14] Lin, J.W. and Mcleod, A.I. (2006) Improved Pena-Rodriguez Portmanteau Test. *Computational Statistics & Data Analysis*, **51**, 1731-1738. <https://doi.org/10.1016/j.csda.2006.06.010>
- [15] Box, E.P.G. (1954) Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *Annals of Mathematical Statistics*, **25**, 290-302. <https://doi.org/10.1214/aoms/1177728786>
- [16] Fisher, T.J. and Gallagher, C.M. (2012) New Weighted Portmanteau Statistics for Time Series Goodness of Fit Testing. *JASA: Journal of the American Statistical Association*, **107**, 777-787. <https://doi.org/10.1080/01621459.2012.688465>
- [17] Davidson, R. and Mackinnon, J.G. (1999) The Size Distortion of Bootstrap Tests. *Econometric Theory*, **15**, 361-376. <https://doi.org/10.1017/S0266466699153040>