

平稳性检验方法的有效性研究

管河山,周 丹^①

(南华大学 经济管理学院,湖南 衡阳 421001)

[摘要] 平稳性检验是时间序列分析的重要研究内容,现有检验方法的性能缺乏系统的比较分析。文章从样本长度的视角研究平稳性检验方法的性能,采用 ADF 检验、PP 检验、KPSS 检验和 LMC 检验四种方法展开实证研究。仿真实验结果表明:时间序列数据长度会对检验方法的准确率产生明显的影响,数据长度较小时检验准确率偏低;数据长度增大时可以提升检验方法的准确率,但仍未能达到 100% 的上限值。当样本长度较小时,这些方法的检验统计量的渐进分布难以满足,因此其实际检验效果值得探究。样本长度是有限的,因此渐进分布检验方式的改进空间有限,新的检验方式值得探究。

[关键词] 时间序列; 平稳性检验; 准确率

[中图分类号] F224 **[文献标识码]** A **[文章编号]** 1673-0755(2016)01-0063-06

时间序列是一类典型的随机数据,基于随机变量的历史和现状来推测其未来需要假设随机变量的历史和现状具有代表性或可延续性,样本时间序列展现了随机变量的历史和现状。如果随机变量基本性态维持不变,则要求样本数据(时间序列)的本质特征仍能延续到未来;统计量(均值、方差、协方差)的取值在未来仍能保持不变,则样本时间序列具有平稳性^[1]。平稳性分析是不同领域的时间序列数据建模的共性问题,时间序列平稳是经典回归分析赖以实施的基本假设。如果数据非平稳,则作为大样本下统计推断基础的“一致性”要求便被破坏,基于非平稳时间序列的预测也就失效。因此,平稳性检验是时间序列分析的一个关键问题。

通常随着时间 t 取值变动,时间序列的期望值、方差和延迟 k 阶自协方差的绝对取值一般是变动的,时间序列平稳性特征将判定为非平稳。因此,实践中时间序列平稳性的检验,大都是从时间序列样本数据取值的角度来分析一些统计量特征,其本质是得到一个平稳性检验的“近似结果”。现有的时间序列数据平稳性检验方法较多,主要包括主观检验方法(如自相关函数图)和客观检验方法(如单位根检验)两类方法,然而这些方法的实际性能如何,尚需要进行一个系统的探讨,这正是本研究的出发点。

从平稳性定义和“近似检验”的分析可知,样本长度将会对检验方法的效果造成明显的影响。刘田^[2]探讨了单位根检验中的样本长度问题,发现单位根检验在分析低维度样本数据时,其实际效果偏差。因此,实证中所采集的数据样本长度太短,则其得到的平稳性分析结果显得不可靠,然而,在实际应用中,仍有大量学者在实证分析时所采集的时间序列样本长度较短^[3,4],因此,其检验结果和后续分析值得怀疑。刘田的研究采用单位根,并针对 AR 系列模型展开,本文从样本长度的视角展开,对 AR (Auto-Regression) 系列、MA (Moving Average) 系列、ARMA (Auto-Regressive and Moving Average) 系列和 ARIMA (Autoregressive Integrated Moving Average) 系列四种模型展开分析,并采用 ADF 检验 (Augmented Dickey-Fuller test)、PP 检验 (Phillips-Perron test)、KPSS 检验 (Kwiatkowski-Phillips-Schmidt-Shin test)、LMC 检验 (Leybourne-McCabe stationarity test) 四种方法展开实证,分析这些平稳性检验方法的实际性能。

一 平稳性检验方法及性能评价

(一) 平稳性检验方法

如何有效地判定时间序列的平稳性,最基本的方法就是平稳性检验。现有方法主要包括主观检验

[收稿日期] 2015-10-28

[基金项目] 教育部青年基金项目“海量金融时间序列数据平稳性检验方法研究”资助(编号:13YJCZH044);南华大学社科基金项目“海量信息视角下的投资决策研究”资助(编号:2012XYB05)

[作者简介] 管河山(1981-),男,湖南祁东人,南华大学经济管理学院副教授,博士(后)。

^①南华大学经济管理学院硕士研究生。

方法和客观检验方法,具体如下:

主观检验方法的典型方式是借助图形判断时间序列的平稳性特征。比如,时间路径图法可以绘制出时间序列的散点图,然后观察散点是否围绕其均值上下波动,如果是,则判断序列为平稳的,否则,判断为非平稳的^[5],该方法受判断者的主观意识影响比较大。其次,时间序列相关函数图为平稳性判定提供了一个有效的途径,这包括自相关函数图(2011)^[6-7]、偏自相关函数图^[8-9]和逆自相关函数图^[10]。对给定的时间序列数据样本,若其自相关函数序列迅速收敛到 0,则判定为平稳的,反之,则判定为非平稳的。偏自相关函数图和逆自相关函数图的应用与自相关函数图相似,这些方法对序列收敛状态的判断带有一定的主观性,因此都属于主观类检验方法。

客观检验方法的典型代表是单位根检验。单位根检验(Dickey-Fuller test, DF 检验)^[11]是平稳性检验中常用到的一种方法。ADF 检验^[12]是一种扩充的 DF 检验,可以解决 DF 检验中对随机项误差白噪声假设的问题。DF 单位根检验和 ADF 单位根检验是通过建立时间序列的自回归方程,引入滞后算子构建特征方程,根据其特征根的绝对值来判断序列的平稳性,其准确性主要受到自回归方程精确性的影响,实践中只能从统计建模角度给出一个最优的模型。Dickey 和 Fuller 使用蒙特卡洛模拟方法计算出了 ADF 检验统计量的临界值表,并用于判断序列的平稳性,其判定结果也是统计意义上的某种近似结果。单位根检验已成为时间序列平稳性检验的主流方法,其后续研究也很多^[13-14]。此外,PP 检验^[15-16]、KPSS 检验^[17-18]、LMC 检验^[19]等检验方法也得到了广泛的研究和应用,这些方法都沿用了统计分布的检验方式。国内外学者的相关研究极大地丰富了时间序列平稳性检验的理论和应用研究。

(二)性能评价

对主观检验方法而言,如果面临大批量的时间序列数据,采取这种人工判定的主观方式进行逐个

分析,其工作量无疑是十分巨大的。而且,这种主观判定的结果将会因人而异,难以得到一致的判定结果,因此,实际应用中,这些主观检验方法常用于对平稳性特征的粗略判定。

对客观检验方法而言,这种借助统计分布临界表的平稳性判定方式,其本质是给出统计意义上的某种近似结果。尽管它们已成为平稳性检验研究的主流方法,然而这些方法的检验统计量的统计分布通常要假定样本长度趋向无穷大,而现实中所采集到的时间序列数据,往往很难满足这些假设;如果现实中采集的时间序列样本长度较短,则这些统计分布的假设条件就不容易满足,此时检验的实际结果值得怀疑。因此,仍需要对客观类检验方法的性能进行系统分析,实证表明其准确率仍不够理想(见本文的实证部分)。

本文主要针对客观类平稳性检验方法进行分析,从时间序列的样本长度视角探讨检验方法的实际性能。

二 实验方法及数据说明

正如前文所述,现有的检验方法较多,本文选取平稳性检验理论研究和应用研究中较为常见的四种检验方法,即 ADF 检验、PP 检验、KPSS 检验、LMC 检验进行分析,重点探讨这些检验方法的性能。对于主观类检验方法,其检验结果因人而异,且主观判定方式处理海量数据时的工作量巨大,因此,本文不做过多探讨。

为了客观评价各检验方法的效果,需要采集一批已知平稳性结果的数据,用于实证测试。为此,本文特生成 6 组非平稳时间序列数据和 6 组平稳时间序列数据进行实验分析^[10],在模型参数的选择方面,尽量使得生成的时间序列难以区分,即难以判定时间序列的平稳性与非平稳性,所以模型参数选择靠近 0 或 1。具体模拟数据的产生方式如表 1 所示:

表 1 模拟的非平稳时间序列

编号	模型	参数	模型表达式
NS1	ARIMA(1,1,0)	$\phi_1 = -0.1$	$y_t = 0.9 * y_{t-1} + 0.1 * y_{t-2} + e_t$
NS2	ARIMA(0,1,0)		$y_t = y_{t-1} + e_t$
NS3	ARIMA(0,1,1)	$\theta_1 = 0.1$	$y_t = y_{t-1} - 0.1 * e_{t-1} + e_t$
NS4	ARIMA(0,1,1)	$\theta_1 = -0.1$	$y_t = y_{t-1} + 0.1 * e_{t-1} + e_t$
NS5	ARIMA(0,1,1)	$\phi_1 = 0.1, \theta_1 = -0.1$	$y_t = 1.1 * y_{t-1} - 0.1 * y_{t-2} + 0.1 * e_{t-1} + e_t$
NS6	ARIMA(1,1,1)	$\phi_1 = 0.05, \theta_1 = -0.05$	$y_t = 1.05 * y_{t-1} - 0.05 * y_{t-2} + 0.05 * e_{t-1} + e_t$

表 2 模拟的平稳时间序列

编号	模型	参数	模型表达式
S1	AR(1)	$\phi_1 = 0.9$	$y_t = 0.9 * y_{t-1} + e_t$
S2	AR(2)	$\phi_1 = 0.95, \phi_2 = -0.1$	$y_t = 0.95 * y_{t-1} - 0.1 * y_{t-2} + e_t$
S3	MA(1)	$\theta_1 = -0.9$	$y_t = 0.95 * e_{t-1} + e_t$
S4	MA(2)	$\theta_1 = -0.95, \theta_2 = -0.1$	$y_t = 0.95 * e_{t-1} + 0.1 * e_{t-2} + e_t$
S5	ARMA(1,1)	$\phi_1 = 0.95, \phi_2 = 0.1$	$y_t = 0.95 * e_{t-1} - 0.1 * e_{t-2} + e_t$
S6	ARMA(1,1)	$\phi_1 = -0.1, \phi_2 = -0.95$	$y_t = -0.1 * y_{t-1} + 0.95 * e_{t-1} + e_t$

其中, e_t 为服从标准正态分布的伪随机函数。模拟实验数据的时间长度分别取 50、100、150、200、250、500 六种长度,每次模拟分别产生 1000 个时间序列数据样本。因此,对每一种长度,都产生 12000 个时间序列数据样本,对六种长度,一共产生 72000 个时间序列数据样本。大量的模拟数据用于测试,一定程度上可以提高实验结果的可靠性。

三 实验结果及分析

采用 Matlab 软件进行分析,具体的实验结果和分析如下。

(一)仿真实验的准确率分析

首先,采用 ADF 检验和 PP 检验进行分析,两种方法的准确率相同,如表 3 所示。

表 3 ADF 检验和 PP 检验的准确率(单位:%)

模 型		时间序列长度					
		L = 50	L = 100	L = 150	L = 200	L = 250	L = 500
非平稳模型	NS1	92.70	93.10	91.20	93.60	93.00	93.50
	NS2	94.30	94.80	96.20	95.00	95.40	94.10
	NS3	92.60	91.90	92.80	92.00	93.00	91.20
	NS4	96.80	96.60	97.00	96.50	96.50	97.30
	NS5	98.40	98.90	98.20	98.90	98.30	98.30
	NS6	97.00	96.60	97.70	96.90	97.00	96.90
平稳模型	S1	32.00	80.50	97.10	99.50	100.00	100.00
	S2	47.80	95.40	99.90	99.90	100.00	100.00
	S3	100.00	100.00	100.00	100.00	100.00	100.00
	S4	100.00	100.00	100.00	100.00	100.00	100.00
	S5	23.40	44.50	70.90	88.60	96.50	100.00
	S6	100.00	100.00	100.00	100.00	100.00	100.00
平均准确率		81.25	91.03	95.08	96.74	97.48	97.61

分析 ADF 检验和 PP 检验的结果,由表 3 可以得出:(1)对 6 组非平稳时间序列数据样本而言,两种检验方法的效果表现很好,检验准确率都达到了 90% 以上,并且效果较为稳定,受时间长度与模型参数影响不大;(2)对 6 组平稳时间序列数据样本而言,两种检验方法的实际准确率有较大的波动,受具体模型的影响明显;对 S1、S2 和 S5 三组模拟数据的检验准确率偏差,且随着样本长度的增加有明显的提升提高,最高可达到 100%;对 S3、S4 和 S6 三组模拟数据的检验准确率非常稳定,都达到了 100%;(3)计算 12 组模型的检验准确率的均值可知,ADF 检验效果受时间序列样本长度的影响显著,样本长度越大,则检验准确率越高。

采用 KPSS 检验进行分析,统计其准确率如表 4 所示。

从表 4 中可以看出:(1)对 6 组非平稳时间序列数据样本而言,KPSS 检验的结果比较理想,无论样本长度取值大小,其检验准确率都超过了 95%;(2)但对 6 组平稳时间序列数据而言,KPSS 检验准确率有很大的波动;对 S1、S2 和 S5 三组模拟数据的检验准确率极差,无论样本长度取值大小,其检验准确率都低于 10%;对 S3、S4 和 S6 三组模拟数据的检验准确率不太理想,其检验准确率位于[58.4%,76.3%]的区间,且受样本长度的影响较小;(3)计算 12 组模型的检验准确率的均值可知,KPSS 检验效果受时间序列样本长度的影响不太明显,随着样本长度的增大,其检验准确率没有明显的提高。

采用 LMC 检验进行分析,统计其准确率如表 5 所示。

表 4 KPSS 检验的准确率(单位:%)

模 型		时间序列长度					
		L = 50	L = 100	L = 150	L = 200	L = 250	L = 500
非平稳模型	NS1	96.5	99.8	99.9	100	100	100
	NS2	97.8	99.8	100	100	100	100
	NS3	96.8	99.8	100	100	100	100
	NS4	97.7	100	100	100	100	100
	NS5	98	100	100	100	100	100
	NS6	98.1	99.9	100	100	100	100
平稳模型	S1	5.4	1.4	0.4	0.1	0	0.1
	S2	7	2.3	1.2	1.2	1	0.4
	S3	70	68.1	71.9	68.1	66.9	67.5
	S4	58.4	64.2	63.8	62.2	61.7	63.1
	S5	4.9	0.4	0	0	0	0
	S6	76.3	72.6	73.2	75	73	71
平均准确率		67.2	67.4	67.5	67.2	66.9	66.8

表 5 LMC 检验准的率率(单位:%)

模 型		时间序列长度					
		L = 50	L = 100	L = 150	L = 200	L = 250	L = 500
非平稳模型	NS1	79.2	85.6	89.4	91.8	94.4	99.4
	NS2	56.8	54	53.1	55.1	51.8	52.7
	NS3	78.7	83.6	91	91.5	94.7	98.3
	NS4	28.1	18.6	11.9	8.3	5.9	2.5
	NS5	12.2	3.4	1.1	0.3	0	0
	NS6	30.1	19	11.8	8.2	7	1.7
平稳模型	S1	34.9	29.1	27.7	23.9	18.7	12.2
	S2	54	58.8	61.4	64	64.8	71.8
	S3	68.6	67.4	67.3	66.4	64.3	67.9
	S4	75	79.6	82.1	85.4	85.9	93.6
	S5	17.8	9.6	5.7	4.7	2.3	0.2
	S6	64.9	58.7	59.7	63.9	61.3	60.3
平均准确率		50.0	47.3	46.9	47	45.9	46.7

从表 5 中可以看出:(1)对 6 组非平稳时间序列数据而言,LMC 检验准确率有很大的波动;检测准确率受模型参数影响很大,NS1、NS2 和 NS3 模型的准确率相对较高,而 NS4、NS5 和 NS6 模型的准确率相对较低,且随时间长度的增加,6 个模型的检验准确率有明显的波动;(2)对 6 组平稳时间序列数据而言,LMC 检验准确率也有很大的波动;检测准确率受模型参数影响很大,NS1 和 NS5 模型的准确率相对较低,而 NS2、NS3、NS4 和 NS6 模型的准确率相对较高;且随时间长度的增加,检验准确率有明显的波动;(3)计算 12 组模型的检验准确率的均值可知,LMC 检验效果受时间序列样本长度的影响不太明显,随着样本长度的增大,其检验准确率没有明显的提高。

(二)仿真实验的检验错误分析

从表 3-表 5 可知,无论是平稳模拟数据还是非平稳模拟数据,四种方法都存在一定概率的检验错

误。如果将单位根误判为平稳过程定义为第一类错误,将平稳过程误判为单位根过程定义为第二类错误。针对各种样本长度,首先计算出四种检验方法的第一类错误的均值,绘制曲线图,如图 1(a)所示;然后,计算出四种检验方法的第二类错误的均值,绘制曲线图,如图 1(b)所示;其中,ADF 检验和 PP 检验两种方法的错误率曲线是重合的。

由图 1 可知:(1)LMC 检验方法的第一类错误率较高,这表明该检验方法有很大概率将非平稳时间序列判定为平稳时间序列,而 ADF 检验、PP 检验和 KPSS 检验三种方法的第一类错误率较低;(2)KPSS 检验和 LMC 检验方法的第二类错误率较高,这表明该检验方法有很大概率将平稳时间序列判定为非平稳时间序列,而 ADF 检验和 PP 检验两种方法的第一类错误率较低;(3)在针对多个仿真模型的实证结果表明,ADF 检验和 PP 检验的检验性能比较理想。

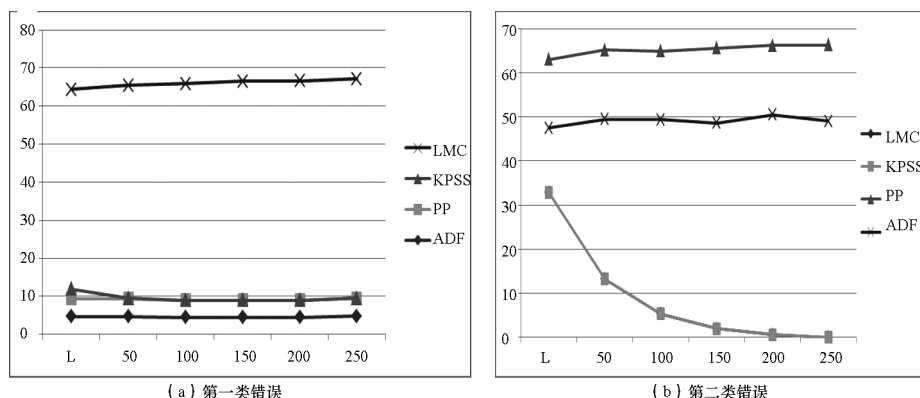


图 1 仿真实验的错误率曲线

四 结 论

平稳性检验是时间序列分析中的一个重要研究内容。主观类平稳性检验方法的实际检验结果会受到个人因素影响而呈现不一致的情况,因此往往用于对平稳性的粗略判定。客观类平稳性检验方法虽然可以给出统计意义上的唯一判定结果,然而这种基于统计分布的近似检验方式,其实际准确率值得深究。本文实证发现:一方面,时间序列样本长度会对客观类检验方法造成影响,这一点与刘田的研究结论是一致的^[2];另一方面,从两类错误分析来看,ADF 检验和 PP 检验是相对较好的判定方法,在仿真实验中呈现较好的性能。然而,对某些模拟数据(模型 S1、S2 和 S5)的检验效果仍很不精确,而且当样本长度较小时,其检验准确率明显较低。综上分析,不可否认,时间序列平稳性检验研究仍然任重道远,采用现有的检验方法进行实践时,采集足够长时间的数据(增加样本长度)是提高检验结果可靠性的必要前提。实际采集到的样本长度是有限的,渐进分布检验方式的改进空间有限,新的检验方式值得探究。

此外,随着计算机技术和数据采集手段的发展,我们可以收集到越来越多的时间序列数据,对这些海量时间序列数据的分析和建模,需要建立一种快速的、高准确率平稳性检验方法进行分析。这将是我們下一步研究的工作重点。

[参考文献]

[1] [美] 伍尔德里奇. 计量经济学导论[M]. 4 版. 费剑平, 译. 北京: 中国人民大学出版社, 2010.

[2] 刘 田. 单位根检验中样本长度的选择[J]. 数理统计与管理, 2013, 32(4): 617-626.

[3] 姚耀军, 和王禅. 农村资金外流的实证分析: 基于结构突变理论团[J]. 数量经济技术经济研究, 2004(s): 28-33.

[4] 陈 龙. 结构性突变的单位根过程—基于中国广义货币的实证[J]. 统计与决策, 2004(11): 12-13.

[5] J D Hamilton. Time Series analysis[J]. Princeton University Press, 1994.

[6] P Galeano. Multivariate Analysis in Vector Time Series[J]. Resenhas, 2000(4): 383-404.

[7] A Alyamkin, S I Eremenko. Pedestrian detection algorithms based on an analysis of the autocorrelation function of a seismic signal[J]. Journal article, 2011, 47(2): 24-129.

[8] W S Cleveland. the Inverse Autocorrelations of a Time Series and Their Applications[J]. Technometrics, 1972(14): 277-293.

[9] O F Durdu. Application of linear stochastic models for drought forecasting in the Büyük Menderes river basin, western Turkey[J]. Stochastic Environmental Research and Risk Assessment, 2010, 24(8): 1145-1162.

[10] C Jorge, C Nuno, P Daniel. A Periodogram-based Metric for Time Series Classification[J]. Computational Statistics & Data Analysis, 2006(50): 2668-684.

[11] D A Dickey. Estimation and Testing of Nonstationary Time Series[D]. Unpublished Ph. D. Thesis, Iowa State University, Ames, 1976.

[12] D A Dickey, W. A Fuller. Distribution of the Estimators for Autoregressive Time Series with a Unit Root[J]. Journal of American Statistical Association, 1979(745): 427-431.

[13] Westerlund Joakim. The effect of recursive detrending on panel unit root tests[J]. Journal of Econometrics, 2015, 185(2): 453-467.

[14] 杨利雄. 平滑结构突变下 Dickey_Fuller 检验的大样本行为[J]. 统计研究, 2013, 30(11): 104-108.

[15] Phillips P, P Perron. Testing for a Unit Root in Time Se-

ries Regression[J]. Biometrika, 1988(75):335-346.

[16] 陈双金. 时间序列单位根检验方法比较[D]. 成都: 电子科技大学硕士学位论文, 2013.

[17] Kwiatkowski D, P C B Phillips P, Schmidt Y Shin. Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root[J]. Journal of Econometrics, 1992(54): 159-178.

[18] 冯 蕾, 聂巧平. 结构突变对 KPSS 检验水平与检验功效的影响——基于有限样本情形的模拟及实证研究[J]. 统计研究, 2009, 26(9): 96-100.

[19] Leybourne S J, B P M McCabe. Modified Stationarity Tests with Data-Dependent Model-Selection Rules[J]. Journal of Business and Economic Statistics, 1999(17): 264-270.

Research on the Validity of Stationary Testing Methods

GUAN He-shan, ZHOU Dan

(University of South China, Hengyang, China, 421001)

Abstract: Stationary test is one of the important parts of time series analysis. Although there exist some testing methods, systemic comparison of their validity is still lack. This paper studies their validity from the aspect of sample length by using ADF test, PP test, KPSS test and LMC test. The experimental results indicate that the data length of time series will dramatically influence the accuracy of these testing methods, which makes testing accuracy lower when it's shorter while improves accuracy through increasing itself but hardly reaches the upper limit of 100%. The progressive distribution of their test statistics is hardly satisfied when the sample length is short, which shows that the actual effect is worth exploring. The length of sample is limited, therefore the improvement of the asymptotic distribution test is restricted, and it's still a significant work to explore new test methods.

Key words: time series; stationary test; accuracy