

211

国家自然科学基金资助项目 教育部人文社会科学博士点基金研究项目
西南财经大学“十五”“211工程”建设项目

靳庭良 著

单位根检验程序的 改进研究

DANWEIGEN JIANYAN CHENGXU DE
GAIJIN YANJIU

西南财经大学出版社

211

国家自然科学基金资助项目 教育部人文社会科学博士点基金研究项目
西南财经大学“十五”“211工程”建设项目

靳庭良 著

单位根检验程序的 改进研究

DANWEIGEN JIANYAN CHENGXU DE
GAIJIN YANJIU

西南财经大学出版社

图书在版编目(CIP)数据

单位根检验程序的改进研究/靳庭良著. —成都:西南财经大学出版社, 2006.8

ISBN 7-81088-543-X

I. 单... II. 靳... III. 时间序列分析—数学模型—应用—金融 IV. F830.41

中国版本图书馆 CIP 数据核字(2006)第 067403 号

单位根检验程序的改进研究

靳庭良 著

责任印制:杨斌

责任编辑:李霞湘

封面设计:杨红鹰

| | |
|-------|--------------------------------------------------------------------------|
| 出版发行: | 西南财经大学出版社(四川省成都市光华村街 55 号) |
| 网 址: | http://press.swufe.edu.cn |
| 电子邮件: | xcpres@mail.sc.cninfo.net |
| 邮政编码: | 610074 |
| 电 话: | 028-87353785 87352368 |
| 印 刷: | 四川五洲彩印有限责任公司 |
| 成品尺寸: | 148mm×210mm |
| 印 张: | 9.375 |
| 字 数: | 230 千字 |
| 版 次: | 2006 年 8 月第 1 版 |
| 印 次: | 2006 年 8 月第 1 次印刷 |
| 印 数: | 1—2000 册 |
| 书 号: | ISBN 7-81088-543-X/F·470 |
| 定 价: | 23.80 元 |

1. 如有印刷、装订等差错,可向本社营销部调换。
2. 版权所有,翻印必究。

前 言

本书是在作者博士学位论文的基础上加以修改而成的,也是“国家自然科学基金项目(70371061)”、“教育部人文社会科学(博士点基金)项目(03JB790011)”和“西南财经大学 211 工程项目课题”的研究成果之一。

单位根检验是建立 ARMA 模型、ARIMA 模型、变量间的协整分析、因果关系检验等的基础。自 Nelson and Plosser(1982)利用 ADF 检验研究了美国名义 GNP 等 14 个历史经济和金融时间序列的平稳性以后,单位根检验业已成为分析经济和金融时间序列变化规律和预测的重要组成部分。因此,单位根检验作为一种特殊的假设检验,其可靠性的研究,以及如何寻求可靠性较高的检验方法或统计量,多年来一直是时间序列分析中的重要课题。本书系统研究了广为应用的单位根检验法——ADF(DF)检验和 PP 检验——的可靠性及检验程序的改进。

关于这些检验稳健性和可靠性的研究,国外学者已做了大量工作,但以往的研究或者缺乏系统性,或者在试验设计上存在一些缺陷,如 Dejong 等(1992)、Kim 和 Schmidt(1993)等,直接由某一常数生成误差项或备择假设的样本序列,这样得到的序列是非(趋势)平稳的,这与单位根检验的初衷不符;Pantula 等(1994)设定初始值为服从特定分布的随机变量,此时无法得到趋势平稳的样本序列。因此他们所得结论都具有一定的局限性。

另外,单位根检验在检验程序上也存在明显的缺陷,致使检验结果通常会给应用者带来误导,这使单位根检验的科学性和可靠性受到质疑。例如,ADF 检验和 PP 检验的理论依据是它们的统计量

与误差项为 i. i. d. 的 DF 统计量,具有相同的极限分布,因此,在有限样本的单位根检验中便可以使用相同的临界值。这里存在的问题是,一方面,极限分布相同的统计量,其有限样本分布可能相差很多;另一方面,ADF 和 PP 统计量极限分布的证明要求滞后截断参数随着样本容量的增加而增加,但速度要慢于样本容量增加的速度。在实证分析中,样本容量是一个确定的数,这一点根本无法验证,因此尽管人们不断改进滞后截断参数的选择方法,但检验水平的扭曲或低势现象无法避免,而且通常表现得相当严重。

原假设的设定是单位根检验的首要问题。现在常用的设定方法,如图示法、Dolado 等(1990)提议的从一般到特殊的检验法等,至少存在以下缺陷:第一,只停留在判断漂移项或趋势项的取舍上,没有充分考虑样本数据生成过程或原假设误差项的设定对单位根检验可靠性的影响。第二,没有考虑所设定原假设过程的可信度。而 Schwert(1987)提出依据样本序列的 ARIMA 模型设定原假设,并以统计量的实际分位数为临界值进行单位根检验,也存在明显的缺陷:一是没有考虑备择假设与原假设表达式的不同对单位根检验势的影响;二是没有考虑一些金融和经济时间序列的波动集聚性引起的 GARCH 效应对单位根检验的影响。

针对单位根检验存在的问题或缺陷,本书从改进模拟试验设计入手,系统研究了 ADF(DF)和 PP 检验的稳健性和可靠性,提出了既有理论依据又能提高检验可靠性的原假设设定策略及单位根检验程序。

本书在附录中给出了笔者开发的具有代表性的 Matlab 模拟程序,这些程序可以举一反三,读者可以方便地应用它们于单位根检验稳健性和可靠性研究及实证分析中。

本书在撰写过程中,得到了西南财经大学博士生导师庞皓教授和我的导师黎实教授的悉心指导,得到了西南财经大学统计学院许多老师和博士生同学的关心和帮助,在此谨对他们致以衷心的感谢。

感谢。

专门研究“单位根检验”的著作,在国内尚无出版先例,由于本人学识肤浅,书中难免有不妥甚至错误之处,恳请读者批评指正。

靳庭良

2006年5月于郑州

摘 要

单位根检验是建立 ARMA 模型、ARIMA 模型、变量间的协整分析、因果关系检验等的基础。自 Nelson 和 Plosser(1982)利用 ADF 检验研究了美国名义 GNP 等 14 个历史经济和金融时间序列的平稳性以后,单位根检验业已成为分析经济和金融时间序列变化规律和预测的重要组成部分。因此,单位根检验作为一种特殊的假设检验,其可靠性的研究以及如何寻求可靠性较高的检验方法或统计量多年来一直是时间序列分析中的重要课题。本书系统研究了广为应用的单位根检验法——ADF(DF)检验和 PP 检验的可靠性及检验程序的改进。

ADF(DF)检验和 PP 检验有比较完整的极限分布理论,但在检验程序和有限样本下的可靠性研究上,它们存在以下缺陷或问题:

(1)模拟试验的初始值问题。一是设定初始值为某常数,直接生成误差项或备择假设过程的样本序列,它们是非(趋势)平稳的,此与 ADF(DF)检验和 PP 检验的原假设为差分平稳,备择假设为(趋势)平稳的设定不相符,如 Dejong 等(1992)、Kim 和 Schmidt(1993)等;另一种是设定初始值为服从特定分布的随机变量,直接生成备择假设的样本序列,如 Pantula(1994)等,这样的设定不能生成趋势平稳的备择假设样本序列。因此,以往有关单位根检验可靠性的研究结论或者存在一定的局限性,或者误导读者。

(2)单位根检验临界值的选择问题。ADF 检验和 PP 检验的理论基础是,在单位根假设成立时,它们的统计量与误差项为 i. i. d. 的相应 DF 统计量具有相同的极限分布,因此在有限样本的单位根检验中便可以使用相同的临界值。这里存在的问题是,一方面,极

限分布相同的统计量,其有限样本分布可能相差很多;另一方面,ADF 和 PP 统计量极限分布的证明要求滞后截断参数随着样本容量的增加而增加,但速度要慢于样本容量增加的速度。在实证分析中,样本容量是一个确定的数,这一点根本无法验证,因此尽管人们不断改进滞后截断参数的选择方法,但检验水平的扭曲或低势现象无法避免,而且通常表现得相当严重。

(3)原假设的设定问题。常用的设定方法,如图示法、Dolado 等(1990)提议的从一般到特殊的检验法等,至少存在以下缺陷:一是只停留在判断漂移项或趋势项的取舍上,没有充分考虑样本数据生成过程或原假设误差项的设定对单位根检验可靠性的影响;二是没有考虑所设定原假设过程的可信度。Schwert(1987)提出依据样本序列的 ARIMA 模型设定原假设,并以统计量的实际分位数为临界值进行单位根检验,也存在明显的缺陷:一是没有考虑备择假设与原假设表达式的不同对单位根检验势的影响;二是没有考虑一些金融和经济时间序列的波动集聚性引起的 GARCH 效应对单位根检验的影响等。

针对单位根检验存在的问题或缺陷,本书从改进模拟试验设计入手,系统研究了 ADF(DF)和 PP 检验的稳健性和可靠性,提出了既有理论依据又能提高检验可靠性的原假设设定策略及单位根检验程序。本书共分八章,各章的主要内容如下:

第一章是绪论。本章首先阐述了单位根检验及其可靠性研究的意义,并对 ADF(DF)检验和 PP 检验的稳健性和可靠性研究状况及影响因素,如单位根过程的误差项性质(包括序列相关性、条件异方差性等)、ADF 检验检验式中的滞后差分项数、PP 检验方差估计中滞后截断参数等,进行评述;其次指出了现有单位根检验在稳健性和可靠性研究中在方法和检验程序上存在的问题,如模拟试验初始值的设定问题、临界值的选择问题和原假设的设定问题等;最后给出了本书所研究的主要问题、研究思路、研究方法以及结构

安排。

第二章给出了生成(趋势)平稳备择假设样本序列的一种一般方法,并据此系统研究了 DF k 、 τ 、F 检验的势以及单位根检验检验式的选择问题。结果表明,在 DF 检验的情形 1 中,对于单位根过程初始值的设定, τ 检验稳健性明显好于 k 检验,并且当 y_0/σ 接近于 0 时, τ 检验的理论势高于 k 检验的理论势,但当 y_0/σ 的绝对值较大时, k 检验比 τ 检验具有较高的理论势;在情形 2 和情形 4 中, k 、 τ 、F 检验的势依次递减,而且与备择假设的初始值、漂移项、趋势项(情形 4)无关;当检验式含有时间趋势项时,所有 DF 检验都只有相当低的势;当检验式与备择假设过程具有相同的表达式时,DF 检验的理论势高于在检验式中增加截距项或趋势项对应检验的理论势。因此,采用何种检验式依赖于单位根原假设是否含有漂移也取决于备择假设的数据生成过程。据此本章最后提出了对传统 DF 检验程序改进的一种方法。

第三章是单位根过程误差项的条件异方差性对 DF 检验影响的研究之一,即当误差项为 GARCH(1,1) - Normal 过程时,研究单位根过程不含有漂移而检验式含有截距项但不含有时间趋势的 DF 检验的稳健性和可靠性。笔者提出了一种模拟生成 GARCH 过程误差项平稳样本序列的一般方法。通过理论分析和随机模拟发现,对于具有 GARCH(1,1) - Normal - errors 的单位根过程采用 DF 统计量进行检验遵循以下规律:①对于任意给定的初始条件方差和条件方差方程的常数项 ω ,当去掉初始生成的数据足够多时,可以得到相当平稳的误差项样本序列,并且 h_0 和 ω 对 DF 统计量分布的影响可以忽略不计;②当方差方程的波动参数较小时,DF 检验是稳健的;③对于给定的检验水平, k 、 τ 、F 检验的理论势依次下降,而且与误差项为 i. i. d. 的情形相比,所有检验的理论势均有一定程度的下降;④对于给定的波动参数和样本容量,条件方差方程的系数和越高,DF 检验的可靠性越差。

第四章是检验式中误差项的条件异方差性对 DF 检验影响的研究之二,即当误差项为 GARCH(1,1) - Normal 过程时,比较研究在各种情形下 DF 检验的稳健性和可靠性以及检验式的选择问题。结果表明,以检验的理论势为标准,当误差项为 GARCH(1,1) - Normal 过程时,DF 检验检验式的选择与误差项为 i. i. d. 情形遵循同样的规律。

第五章是在第三章研究方法的基础上研究扩展的 GARCH 过程,包括条件分布为较正态分布肥尾的 GARCH 误差过程和 EGARCH - Normal 误差过程,对 DF 检验稳健性和可靠性的影响。结果表明,误差项为 GARCH - $t(v)$ 过程与 GARCH - Normal 过程的单位根 DF 检验的稳健性和理论势具有非常类似的变化规律,但它们的实际检验水平和理论势在量上存在着较大的差异,而且条件 $t(v)$ 分布的自由度 v 对 DF 检验有明显影响;误差项对条件方差影响的非对称性对 DF 检验的势也产生一定的影响。因此,在实证分析中,除需要检验样本序列的条件方差时变性外,有时还需要进一步识别条件分布的类型和方差方程中的非对称性,以使单位根检验具有更高的可靠性。

第六章研究了误差项分别为 AR(p) 过程、MA(1) 过程和存在条件异方差性的 AR(p) 过程时,ADF 检验的稳健性和理论势。对一组具有代表性的单位根过程的研究结果表明,无论采用何种方法确定检验式中滞后截断参数 l ,传统 ADF 检验通常都存在较严重的检验水平扭曲问题;GARCH 效应的存在对取得最高理论势的 l 的选择影响不大,但会导致理论势的下降;当误差项为高度负相关的 MA(1) 过程时,滞后截断参数较高的 τ 检验具有最高的理论势,除此之外, l 取值较小的 k 检验具有最高的理论势。最后,在原假设已设定的基础上,依据检验的理论势,提出了一种提高单位根检验可靠性的 ADF(DF) 检验程序。

第七章研究了误差项分别为 AR(p) 过程、MA(1) 过程和存在

条件异方差性的 $AR(p)$ 过程时 PP 检验的稳健性和可靠性。对一组具有代表性的单位根过程的研究结果表明,以 DF 临界值为检验标准,PP 检验通常存在严重的检验水平扭曲;当误差项为高度负相关的 $MA(1)$ 过程时,滞后截断参数 l 较高的 τ 检验具有最高的理论势,除此之外,选择 l 较小的 k 检验具有最高的理论势;GARCH 效应的存在对取得最高理论势的选择影响不大,但会导致理论势的下降;以理论势为可靠性的比较标准,ADF 检验和 PP 检验可靠性的高低依赖于原假设的数据生成过程和样本容量。

第八章探讨了单位根检验原假设的设定及其应用。研究结果表明,单位根原假设的不同设定对检验结果产生很大的影响。以样本序列的 ARMA(或 ARMA-GARCH 类)模型为基础设定原假设和直接以差分模型设定原假设两种方法为基础,提出了一种单位根检验原假设的设定策略:首先以差分序列模型为原假设检验;如果检验拒绝原假设,则认为原序列为平稳或趋势平稳的,否则,在原序列模型的基础上,重新建立原假设进行检验;如果检验拒绝原假设,则认为原序列为平稳或趋势平稳的;否则,则认为原序列的数据生成过程为单位根过程。最后利用改进的单位根检验程序研究了我国国内生产总值的单整性和沪深股指日、周、月收益率序列的平稳性。

本书的创新之处主要体现在以下几个方面:

(1)通过理论分析和模拟试验,指出了已有单位根检验的有限样本性质研究在试验设计上的缺陷,给出了由某一初始值生成误差过程的平稳样本序列或备择假设过程(趋势)平稳样本序列的一般方法,同时在理论上对该方法的合理性进行了论证,纠正了已有研究所得到的可能产生误导的一些结论。

(2)在以统计量的实际分位数为临界值,保证实际检验水平忠实于名义检验水平的理念下,通过模拟检验的势,系统研究了 ADF(DF)检验和 PP 检验的可靠性及其与滞后截断参数和检验式之间的关系;对于一些典型误差项情形,得到了使单位根检验具有较高

可靠性的选择滞后截断参数、统计量和检验式的一般性规律,以及 ADF(DF)和 PP 检验(含各种统计量)适用范围。

(3)剖析了单位根检验原假设设定和检验程序中存在的问题;以样本序列的 ARMA(或 ARMA - GARCH 类)模型为基础设定原假设和直接以差分模型设定原假设两种方法为基础,提出了一种既有一定的理论根据又能提高单位根检验可靠性的原假设设定策略及综合检验程序。

Abstract

Distinguishing between difference stationary and stationary or trend stationary has been become a common starting point, because the characteristics of a non – stationary time series change over time. This equals to answer whether contains a unit root in data generating process of sample series. Since Nelson and Plosser (1982) studied the stationarity of 14 historical economic time series of USA using ADF test, unit root tests have been become an important part in empirically analyzing or forecasting economic and financial time series' changing laws, in which random walk test plays an important role in verifying the theories of rational expectations. On the other hand unit root test is also the theoretical foundation for establishing ARMA models or ARIMA models, cointegration analyzing, and casualty testing etc. so how to choose more reliable unit root testing procedure or statistic is an important problem in time series analysis. This dissertation studies mainly the reliabilities of the most popular ADF (DF) tests and PP tests.

The reliability of a unit root test depends on an appropriate trade off between type – I error and type – II error. Given one of the probabilities of making the two errors, the smaller the other is, it means the more reliable the unit root test is. In practice, the probability of making type – I error is usually given, the test with higher power is chosen. On the other hand, the robustness is also an important standard for evaluating a unit root test, which is the degree of actual size biasing from nominal size, the smaller the degree is, the better the robustness is.

Since ADF (DF) tests and PP tests were introduced, the researches about the reliabilities and robustness of them have not been disconnected. But there exist at least following drawbacks in the past researches:

(1) Studying unit root tests' reliability using Monte Carlo simulation experiment, it is an important step to determinate the initial value for generating sample series. Because Dejong et al, Kim and Schmidt and Ling et al etc supposed the initial value as some constant, the sample series of the errors of regressive models and of alternative hypothesis processes are non-stationary or non trend stationary, which do not accord with the original intentions of ADF (DF) tests and PP tests. Pantula et al supposed the initial value as a random variable obeying some distribution, which would obtain stationary sample series for alternative hypothesis processes not including time trend. However, for alternative hypothesis process with trend, this method is impracticable.

(2) ADF (DF) tests and PP tests depend on the idea that under unit root hypothesis, ADF statistics and PP statistics have the same limit distributions with relevant DF statistics under i. i. d. errors, so the same critical value may be used for big sample. There exist two drawbacks in these tests' applications. (a) The size of sample volume is relative, so real "big sample" does not exist in reality. (b) The proofs of the limit distributions of ADF statistics and PP statistics request that the truncation lag length increases as the sample volume increases, which would not been verified for a given sample. So although the methods for the selection of the truncation lag have been improved from different respects, there usually exist size distortions, which sometime are very serious, as ADF tests and PP tests are performed.

(3) The different setting up of the null hypothesis make often great effect on the results of ADF (DF) tests and PP tests, which has not

been paid enough attention when they are performed using Fuller critical values as the tests' standards. The consequences would be result in serious size distortions usually.

These drawbacks affect the reliabilities of ADF (DF) tests and PP tests directly. So it's significant both in theory and in practice that they are studied objectively and systematically. This dissertation studies these tests' reliabilities through correcting the simulation experiment designs, furthermore makes some improvements for the procedures of these tests. Finally combining with empirically analyzing, a new test strategy is proposed. Based on this train of thoughts, this dissertation devotes eight chapters to study the related issues respectively.

Chapter one is an introduction of this dissertation. First it expounds the significance for studying the reliabilities of ADF (DF) tests and PP tests and analyzes the factors affecting these reliabilities, including the characters of the innovation of the data – generating process, the truncation lag in ADF tests' regressive models and in PP tests' variance estimations, etc. Secondly it points out some drawbacks in experiments and in test procedures which exist in the past researches for the robustness and reliabilities of ADF (DF) tests and PP tests. Finally it indicates the main issues that will be resolved, the train of thoughts and the methods that the article follows.

Chapter two introduces a general method for generating stationary sample series of the alternative hypothesis processes in unit tests' simulation experiments, further studies the powers of DF k test, τ test and F test in all cases, and how to choose DF tests' regressive models, getting the following findings. (1) In case 1, τ test' theoretical powers are higher than k test's as ρ closes to zero, otherwise the result is contrary, in which power obtained by simulating is called as theoretical

power if the critical value is the actual percentile of the statistic under the null hypothesis. (2) in case 2 and in case 4, the power of k test, τ test and F test decrease in turn, which has nothing to do with the initial values, drifts and linear trends (in case 4) of the alternative hypothesis process, but all tests have very low powers in case 4. (3) The theoretical powers of DF tests are higher when the autoregressive models have the same expressions with alternative processes than when they have redundant drifts or trends. So which autoregressive model is chose depends on both the unit root null hypothesis process and the alternative hypothesis process. At last according above conclusions, the author corrects the traditional test procedure of DF tests.

Chapter three points out some drawbacks in Kim and Schmidt's experiment design, studies the robustness and reliabilities of DF tests in case 2 for unit root processes with $GARCH(1,1)$ - Normal - errors and finds several basic laws by analyzing in theory and Monte Carlo simulating. (1) For any given initial conditional variance and the constant ω of the conditional variance equation, when enough data generated initially are dropped, the effect of ω on the DF statistics' distributions may be ignored. (2) The DF tests are robust as the volatility parameter is small. (3) The test through k statistic is more reliable than through statistic and through F statistic. (4) The higher the integratedness of $GARCH$ - error is, the less reliable the DF tests behave.

According to the method introduced in chapter three, Chapter four studies the robustness and reliabilities of DF tests in all other cases and the problem about the choice of the autoregressive models of DF tests for unit root processes with $GARCH(1,1)$ - Normal - errors. The results indicate the same laws are followed in choosing autoregressive models of DF tests by the theoretical powers as the cases in which the errors are

i. i. d. , i. e. the GARCH effects have not remarkable influences on the choice of regressive models of DF tests.

Chapter five studies the extended GARCH processes' effects on the robustness and reliabilities of DF tests, including GARCH – errors with conditional t distributions and EGARCH – Normal – errors. The results indicate (1) when the errors are GARCH – $t(v)$ processes, the actual sizes and the theoretical powers of DF tests follow the similar changing laws as the cases in which the errors are GARCH – Normal processes, but differ widely in amount. (2) The asymmetry of the disturbances' effects on conditional variance has certain extent influence on the actual size and the theoretical power. So in empirically analyzing, besides testing for conditional heteroskedasticity of sample series, it is necessary testing the asymmetry in conditional variance equation so that enhancing unit root tests' reliability.

Chapter six is devoted to the research of the robustness and reliabilities of ADF tests in case 2 and the relations between them and the truncation lag, including three cases: $AR(p)$ – errors, $AR(p)$ – errors with conditional heteroskedasticity and $MA(1)$ – errors. The results indicate no matter what methods are adopted for the selection of the truncation lag, ADF tests using Dickey – Fuller critical values are all likely to result in size distortions and/or power loss; for the $MA(1)$ – errors which are highly negatively correlated, the ADF test with some larger truncation lag has the highest theoretical power, otherwise ADF k test with smaller truncation lag has the highest theoretical power; On basis of above research results, a new ADF (DF) testing procedure is given.

Chapter seven studies the robustness and reliabilities of PP tests in case 2 and the relations between them and lag truncations, including three cases: $AR(p)$ – errors, $AR(p)$ – errors with conditional het-