学校编码：10384　　　　分类号______密级______

学号：19020071154215　　　　UDC______

厦门大学

硕　士　学　位　论　文

自助单位根检验

Bootstrap Unit Root Test

方　书　英

指导教师姓名：　黄　荣　坦　副教授

专　业　名　称：　概率论与数理统计

论文提交日期：　2010 年　　月

论文答辩时间：　2010 年　　月

学位授予日期：　2010 年　　月

答辩委员会主席：__________

评　　阅　　人：__________

2010 年　　月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　　　　　　　　　　　）课题（组）的研究成果，获得（　　　　　　　　）课题（组）经费或实验室的资助，在（　　　　　　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

摘要

近些年来, 自助单位根检验在很多方面得到应用, 甚至在非线性时间序列领域上. 本文介绍了单位根检验和自助法的基本知识. 主要讨论自助DF检验, 自助ADF检验的一般方法和理论知识. 总结了当前自助单位根检验的最新成果, 主要是剔除趋势的自助单位根检验. 并在理论的基础上, 利用蒙特卡罗模拟方法做实证分析, 比较了DF检验和自助DF检验, ADF检验和自助ADF检验在拒绝概率和检验功效上的差别. 用单位根检验和自助单位根检验分析了中国的国债利率和1978年到2009的GDP数据, 从概率角度验证了自助单位根检验的优越性.

关键词:自助单位根检验; 自助DF检验; 自助ADF检验.

Abstract

In recent years the bootstrap unit root tests have become increasingly popular and they have been applied to a wide range of topics, including nonstationary time series. In this paper we review the basic knowledge of the unit root tests and the bootstrap. We mainly discuss general methods and theories of the bootstrap DF tests and the bootstrap ADF tests. We summarize the latest achievements of the bootstrap unit root tests, including detrending bootstrap unit root tests. On the basis of theory, we use the method of Monte-Carlo Simulation to do lots of empirical analysis. We compare the differences between the DF tests and the bootstrap DF tests and differences between the ADF tests and the bootstrap ADF tests on the rejection probabilities and the power of test. We use the unit root tests and the bootstrap unit root tests to analyze the national debt rates and the Chinese GDP dates during the period from 1978 to 2009. We prove the superiority of the bootstrap from the angle of probability.

Key words: Bootstrap Unit Root Tests; Bootstrap DF Test; Bootstrap ADF Test.

目录

Contents

第一章 引言

§1.1 研究背景

单位根检验时间序列的单位根研究是时间序列分析的一个热点问题. 时间序列矩特性的时变行为实际上反映了时间序列的非平稳性质. 对非平稳时间序列的处理方法一般是将其转变为平稳序列, 这样就可以应用有关平稳时间序列的方法来进行相应得研究. 对时间序列单位根的检验就是对时间序列平稳性的检验, 非平稳时间序列如果存在单位根, 则一般可以通过差分的方法来消除单位根, 得到平稳序列. 对于存在单位根的时间序列, 一般都显示出明显的记忆性和波动的持续性, 因此单位根检验是讨论有关协整关系存在性检验和序列波动持续性讨论的基础. 在实际的经济和金融领域有广泛的应用. 一般意义上自助法就是按重抽样方法多次抽取样本, 进行检验的方法. 但通常在独立同分布的样本中重新抽样. 单位根检验是对时间序列的检验, 因此如何对时间序列的样本进行重新抽样是自助单位根检验中必须要解决的问题. 一般有两种方法, 一种是分块抽样方法, 另一种是seiv方法. 在后文中将详细的介绍.

早在上世纪八九十年代, 人们就开始了对自助单位根的研究, 并且取得了可喜的成就. R.L.Taylor(1991)[1]讨论了一阶自回归模型的自助法. Juan Romo(1993)[2]探讨了一阶自回归模型的自助单位根检验. 后来有大量的数学家对自助单位根检验作出积极的贡献, Joon Y.Park(2000)[3], (2003)[4]对于自助单位根的渐近性质给出了理论上的证明. A.M.Robert Taylor(2008)[5]讨论了带有非平稳波动率时间序列的单位根检验.

当前关于自助单位根检验已经有丰硕的成果, 包括Giuseppe Cavaliere(2009)[6]自助M检验, Dilem Yildirim(2009)[7]非线性门限模型的自助单位根检验, Stephan Smeekes(2009)[8]剔除趋势的自助单位根根检验, 以及Nikolay Gospodinov(2009)[9]误

差项是$GARCH(1,1)$模型的自助单位根检验.

自助单位根针对小样本的时间序列, 有着很大的优势. 通过自助法的重抽样得到大量的样本, 从根本上弥补了小样本单位根检验在渐近性质上不足, 在当前的金融时间序列分析上, 有着积极广泛的应用. 例如Xiao Qin(2006[10])利用自助马尔科夫单位根检验, 分析了香港和韩国的房地产市场的泡沫.

自助法需要经过多次抽样, 计算繁杂. 随着计算机技术的迅速发展, 人们已经可以熟练的利用计算机软件进行自助抽样, 进行分析. 随着当前世界经济的繁荣, 金融的地位越来越重要. 因此, 自助单位根检验在金融时间序列领域上将得到越来越广泛的运用.

§1.2 本文的主要工作

总结了前人的工作成果, 主要介绍自助DF检验, 自助ADF检验的方法和理论. 以及当前自助单位根检验的最新成果, 主要是剔除趋势的自助单位根检验.

与现有的文献相比, 本文的主要工作在于下面两点：

第一, 在前人的基础上, 总结了自助单位根检验的理论知识, 系统地阐述了自助单位根检验的一般性方法和步骤.

第二, 同时对单位根和自助单位根检验的拒绝概率和检验功效进行比较, 定量的对比了两种方法的差异性, 说明了自助法的优势. 用单位根检验和自助单位根检验两种方法分析了中国历年的国债利率和GDP数据, 验证了自助单位根检验的合理性.

§1.3 本文的结构

本文的结构安排如下：

第一章, 介绍研究背景, 简单说明本文所做的主要工作, 给出本文的结构和轮廓.

第二章, 具体的理论和研究方法. 首先, 介绍自助法的基本内容，以及单位根检验中常见的几种检验, 包括DF检验, ADF检验, Phillips-Perron检验, M检验. 其次, 总结了自助DF检验和自助ADF检验的理论和步骤. 最后, 介绍当前最新的研究成果中的剔除趋势的单位根检验.

第三章, 本文利用R软件, 蒙特卡罗模拟方法和自助方法进行实证分析. 通过模拟对DF检验和自助DF检验在不同显著性水平下拒绝概率和检验功效进行比较, 以及ADF检验和自助ADF检验在0.05的显著性水平下, 对于差分项参数不同的序列的拒绝概率和检验功效进行比较. 最后, 选取了2001年到2008年四十四期的国债利率数据和1978年至2009年中国的GDP数据, 利用带有趋势的ADF检验和自助ADF检验, 分析了它们各自的单位根检验的结果.

第四章, 总结和展望.

第二章 研究理论及方法

§2.1 预备知识

2.1.1 bootstrap方法[11]

2.1.1.1 非参数bootstrap方法

设总体分布F未知, 但已经有一个容量为n的来自分布F的数据样本, 自这一样本按放回抽样的方法抽取一个容量为n的样本, 这种样本称为bootstrap样本或称为自助样本. 相继地, 独立地自原始样本中取得很多个bootstrap样本, 利用这些样本对总体F进行统计推断. 这种方法称为非参数bootstrap方法, 又称自助法. 这一方法可以用于当人们对总体知之甚少的情况, 它是近代统计中的一种用于数据处理的重要使用方法. 这种方法的实现需要在计算机上作大量的计算, 随着计算机威力的增长, 它已成为一种流行的方法.

bootstrap方法是Efron在20世纪70年代后期建立的.

(一) 估计量的标准误差的bootstrap估计

在估计总体未知参数θ时, 人们不但要给去θ的估计$\widehat{\theta}$, 还需指出这一估计$\widehat{\theta}$的精度, 通常我们用估计量$\widehat{\theta}$ 的标准差$\sqrt{D(\widehat{\theta})}$来度量估计的精度. 估计量$\widehat{\theta}$的标准差$\widehat{\widehat{\theta}} = \sqrt{D(\widehat{\theta})}$也称为估计量$\widehat{\theta}$ 的标准误差.

设$X_1, X_2, \cdots, X_n$是来自以$F(x)$为分布函数的总体的样本, θ是我们感兴趣的未知参数, 用$\widehat{\theta} = \widehat{\theta}(X_1, X_2, \cdots, X_n)$作为$\theta$的估计量. 在应用中$\widehat{\theta}$的抽样分布常是很难处理的. 这样, $\sqrt{D(\widehat{\theta})}$常没有一个简单的表达式, 不过我们可以用计算机模拟的方法来求得$\sqrt{D(\widehat{\theta})}$的估计. 为此, 自$F$产生很多容量为$B$的样本, 对于每一个样

本计算$\widehat{\theta}$的值, 得$\widehat{\theta_1}$，$\widehat{\theta_2}$, $\cdots\cdots$，$\widehat{\theta_n}$ 则$\sqrt{D(\widehat{\theta})}$可以用

$$\widehat{\sigma}=\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\widehat{\theta}_i-\overline{\theta})^2}, \qquad (2-1)$$

来估计, 其中$\overline{\theta}=\frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i$. 然后$F$常常是未知的, 这样就无法产生模拟样本, 不能得到$(2-1)$式的结果, 需要另外的方法.

现在设分布F未知, $x_1,x_2,\cdots,x_n$是来自F的样本值, F_n是相应的经验分布函数, 当n很大时, F_n接近F(见格里汶科定理), 我们用F_n代替上一段中的F, 在F_n中抽样. 在F_n抽样, 就是在原始样本$x_1,x_2,\cdots,x_n$中每次随机地取一个个体做放回抽样. 如此得到一个容量为n的样本x_1^*, $x_2^*,\cdots,x_n^*$. 这就是第一段所说的bootstrap样本. 用bootstrap样本按上一段计算估计$\widehat{\theta}(x_1,x_2,\cdots,x_n)$那样求出$\theta$的估计$\widehat{\theta}^*=\widehat{\theta}(x_1^*,x_2^*,\cdots,x_n^*)$. 估计$\widehat{\theta}^*$称为$\theta$的bootstrap估计.

相继地、独立地抽得B个bootstrap样本, 并且以这些样本分别求出θ的相应bootstrap估计如下:

bootstrap样本1 $x_1^{*_1},x_2^{*_1},\cdots,x_n^{*_1}$, bootstrap估计$\widehat{\theta}_1^*$

bootstrap样本2 $x_1^{*_2},x_2^{*_2},\cdots,x_n^{*_2}$, bootstrap估计$\widehat{\theta}_2^*$

$\vdots \qquad \vdots \qquad \vdots$

bootstrap样本B $x_1^{*_B},x_2^{*_B},\cdots,x_n^{*_B}$, bootstrap估计$\widehat{\theta}_B^*$

则$\widehat{\theta}$的标准误差$\sqrt{D(\widehat{\theta})}$, 就以

$$\widehat{\sigma}=\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\widehat{\theta}_i^*-\overline{\theta}^*)^2}, \qquad (2-2)$$

来估计. 其中$\overline{\theta}^*=\frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i^*$, $(2-2)$式就是$\sqrt{D(\widehat{\theta})}$的bootstrap估计.

综上所述得到求$\sqrt{D(\widehat{\theta})}$的bootstrap估计的步骤是:

1°, 自原始数据样本$x=(x_1,x_2,\cdots,x_n)$按放回抽样的方法, 抽得容量为n的样本$x=(x_1^*,x_2^*,\cdots,x_n^*)$(称为bootstrap样本).

2°, 相继地、独立地求出B个($B\geq 1000$)容量为n的bootstrap样本, $x^{*i}=(x_1^{*i},x_2^{*i},\cdots,x_n^{*i}),i=1,2,\cdots,B$. 对于第$i$个bootstrap样本, 我们计算出$\widehat{\theta}_i^*=\widehat{\theta}(x_1^{*i},x_2^{*i},\cdots,x_n^{*i}),i=1,2,\cdots,B$. ($\widehat{\theta}_i^*$称为$\theta$的第$i$个bootstrap估计)

3°, 计算

$\widehat{\sigma}=\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\widehat{\theta}_i^*-\overline{\theta}^*)^2}$,其中$\overline{\theta}^*=\frac{1}{B}\sum_{i=1}^{B}\widehat{\theta}_i^*$.

(二) 估计量的均方误差及偏差的bootstrap估计

设$X=(X_1,X_2,\cdots,X_n)$是来自总体F的样本, F未知, $R=R(X)$是感兴趣的随机变量, 它依赖于样本X. 假设我们希望去估计R的分布的某些特征. 例如R的数学期望$E_F(R)$, 就可以按照上面所说的三个步骤1°, 2°, 3°进行, 只是在2°中对于第i个bootstrap样本$x_i^*=x_1^{*i},x_2^{*i},\cdots,x_n^{*i}$, 计算$R_i^*=R(x_i^*)$代替计算$\theta_i^*$, 且在3° 中计算感兴趣的$R$的特征, 例如如果希望估计$E_F(R)$, 就计算

$$E_*(R^*)=\frac{1}{B}\sum_{i=1}^{B}R_i^*.$$

(三) bootstrap置信区间

下面介绍一种求未知参数θ的bootstrap置信区间的方法.

设$X=(X_1,X_2,\cdots,X_n)$是来自总体F容量为n的样本, $x=(x_1,x_2,\cdots,x_n)$是一个已知的样本值, F中含有未知参数θ, $\widehat{\theta}=\widehat{\theta}(X_1,X_2,\cdots,X_n)$是$\theta$的估计量, 现在来求$\theta$的置信水平为$1-\alpha$的置信区间.

相继地、独立地从样本$x=(x_1,x_2,\cdots,x_n)$中抽取B个容量为n的bootstrap样本, 对于每个bootstrap样本求出θ的bootstrap估计: $\widehat{\theta}_1^*$, $\widehat{\theta}_2^*$, $\cdots$, $\widehat{\theta}_B^*$. 将它们自小到大排序, 得

$$\widehat{\theta}_{(1)}^*\leq\widehat{\theta}_{(2)}^*\leq\cdots\leq\widehat{\theta}_{(B)}^*.$$

取$R(X)=\widehat{\theta}$, 用对应的$R(X^*)=\widehat{\theta}^*$的分布作为$R(X)$的分布的近似, 求出$R(X^*)$的分布的近似分位数$\widehat{\theta}^*_{\frac{\alpha}{2}}$和$\widehat{\theta}^*_{1-\frac{\alpha}{2}}$, 使

$$P(\widehat{\theta}^*_{\frac{\alpha}{2}}<\widehat{\theta}^*<\widehat{\theta}^*_{1-\frac{\alpha}{2}})=1-\alpha. \qquad (2-3)$$

记$k_1=[B*\frac{\alpha}{2}]$, $k_2=[B*(1-\frac{\alpha}{2})]$, 在(2—3)式中以$\widehat{\theta}^*_{k_1}$和$\widehat{\theta}^*_{k_2}$分别作为分位数$\widehat{\theta}^*_{\frac{\alpha}{2}}$和$\widehat{\theta}^*_{1-\frac{\alpha}{2}}$的估计, 得到近似等式

$$P(\widehat{\theta}^*_{k_1}<\widehat{\theta}^*<\widehat{\theta}^*_{k_2})=1-\alpha,$$

于是由上式就得到θ的置信水平为$1-\alpha$的近似置信区间$(\widehat{\theta}^*_{k_1},\widehat{\theta}^*_{k_2})$, 这一区间称为$\theta$的置信水平$1-\alpha$的bootstrap置信区间. 这种求置信区间的方法称为分位数法.

(四) 用bootstrap-t法求均值μ的bootstrap的置信区间

设$X=(X_1,X_2,\cdots,X_n)$是来自总体F容量为n的样本, $x=(x_1,x_2,\cdots,x_n)$是一个已知的样本值, 均值μ和方差σ^2均为未知参数, 我们要利用样本值x来估计μ.

考虑函数

$$t=\frac{\overline{x}-\mu}{S/\sqrt{n}}, \qquad (2-4)$$

假设总体F具有正态分布, 此时t的分布与参数μ无关, 它是一个枢轴量, 而且有$t\sim t(n-1)$, 利用枢轴量t, 就能求得μ的置信区间. 现在, 总体F不具有正态分布, 但可证t仍是一个枢轴量, 然而t的分布就不是$t(n-1)$分布, 下面我们用bootstrap方法来求μ的近似置信区间.

以原始样本$x=(x_1,x_2,\cdots,x_n)$的样本均值$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$作为$\mu$的估计, 考虑与$t$相应的枢轴量

$$W^*=\frac{\overline{X}^*-\overline{x}}{S^*/\sqrt{n}}, \qquad (2-5)$$

此处$\overline{X}^*$, S^*分别为$\overline{X}$, S相应的bootstrap样本均值与样本标准差. 用W^*的分布近似t的分布, 求出W^* 的近似分位数$w^*_{\frac{\alpha}{2}}$, $w^*_{1-\frac{\alpha}{2}}$, 使

$$P\{w^*_{\frac{\alpha}{2}}<\frac{\overline{X}^*-\overline{x}}{S^*/\sqrt{n}}<w^*_{1-\frac{\alpha}{2}}\}=1-\alpha, \tag{2-6}$$

于是近似有

$$P\{w^*_{\frac{\alpha}{2}}<\frac{\overline{X}^*-\overline{x}}{S^*/\sqrt{n}}<w^*_{1-\frac{\alpha}{2}}\}=1-\alpha,$$

或即

$$P\{\overline{X}-w^*_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}<\mu<\overline{X}-w^*_{\frac{\alpha}{2}}\}=1-\alpha. \tag{2-7}$$

将W^*的B个bootstrap值自小到大排序

$$w^*_{(1)}\leq w^*_{(2)}\leq\cdots\leq w^*_{(B)}.$$

记$k_1=[B*\frac{\alpha}{2}]$,$k_2=[B*(1-\frac{\alpha}{2})]$, 在(2—7)式中以$w^*_{(k_1)}$和$w^*_{(k_2)}$分别作为分位数$w^*_{\frac{\alpha}{2}}$, $w^*_{1-\frac{\alpha}{2}}$的估计, 由(2—7)式得到近似等式

$$P\{\overline{X}-w^*_{(k_2)}\frac{S}{\sqrt{n}}<\mu<\overline{X}-w^*_{(k_1)}\frac{S}{\sqrt{n}}\}, \tag{2-8}$$

由(2—8)式得到μ的置信水平为$1-\alpha$的bootstrap置信区间

$$(\overline{X}-w^*_{(k_2)}\frac{S}{\sqrt{n}},\ \overline{X}-w^*_{(k_1)}\frac{S}{\sqrt{n}}).$$

这一方法称为bootstrap-t法.

用非参数bootstrap方法, 没有假设所研究的总体的分布函数F的形式, bootstrap样本是来自已知的数据(原始样本), 所以称为非参数bootstrap方法.

2.1.1.2 参数bootstrap方法

假设所研究的总体分布的分布函数$F(x;\beta)$的形式已知, 但其中包含未知参数β(β可以是向量). 现在已知有一个来自$F(x;\beta)$的样本

$$X_1, X_2, \cdots, X_n.$$

利用这一样本求出β的最大似然估计$\widehat{\beta}$. 在$F(x;\beta)$中以$\widehat{\beta}$代替β得到$F(x;\widehat{\beta})$, 接着在$F(x;\widehat{\beta})$中产生容量为n的样本

$$X_1^*, X_2^*, \cdots, X_n^* \sim F(x;\widehat{\beta}).$$

这种样本可以产生很多个. 例如产生B个($B \geq 1000$). 就可以利用这些样本对总体进行统计推断, 其做法与非参数bootstrap方法一样, 这种方法称为参数bootstrap法.

2.1.2 单位根检验

单位根过程即$I(1)$过程, $I(1)$过程的一阶差分为平稳过程, 区别$I(0)$和$I(1)$过程的一般的检验都以$I(1)$为原假设, 由于没有研究人员普遍接受的具有较好的有限样本性质的以$I(0)$为原假设的检验, 这里主要讨论原假设为$I(1)$的检验. 常用的检验有DF检验, ADF检验, Phillips检验, M检验等等.

2.1.2.1 DF 检验[12]

(一) 没有截距和时间趋势的情形

数据生成过程(DGP):

$$y_t = y_{t-1} + u_t, \qquad (2-9)$$

其中u_t是$i.i.d.$均值为0, 方差为σ^2.