

文章编号: 1000-5013(2009)03-0338-05

# 中国股市长记忆检验的滑动分块自助法仿真

苏桔芳, 胡日东, 陈家干

(华侨大学 商学院, 福建 泉州 362021)

**摘要:** 提出基于滑动分块自助的 GPH (Geweke, Porter-Hudak) 检验方法, 并检验沪、深股市的各种收益率序列的长记忆性. 结果表明, 沪、深两市的日指数收益率序列是一个  $I(0)$  过程而非长记忆过程. 然而, 沪、深两市的绝对值收益率和收益率平方序列却是一个分数差分过程, 沪市的绝对值收益率和收益率平方序列的分数差分值大约为 0.30, 深市的绝对值收益率和收益率平方序列分数差分值大约为 0.35. 即深市的长记忆性强于沪市, 同时也说明上海股市的运行效率要高于深圳股市.

**关键词:** 长记忆; 分整自回归移动平均模型; 滑动分块自助法; GPH 检验法; 中国股市

**中图分类号:** O 212.1; F 224.0

**文献标识码:** A

在能体现“长记忆”的分整自回归移动平均 (ARFIMA) 模型中, 描述长记忆特征的是其分数差分参数  $d$ , 当  $0 < d < 0.5$  时, ARFIMA 模型表现出显著的长记忆的特征. 对于参数  $d$  的估计方法, 学者至今提出了诸如聚合方差法、聚合序列绝对值方法、重标刻度极差 (R/S) 法、GHP (Geweke, Porter-Hudak) 法等众多方法. 前两种方法由于需要大量的观测样本, 在实践中较为少用, 因此, 后两者成为重要的分析方法. 然而, 这些方法的结论只是渐进成立或者是建立在严格假设基础之上的, 在实际应用中得到的结论常常难以令人信服<sup>[1]</sup>. 因此, 研究参数  $d$  在不受分布约束或者样本容量大小约束的条件的性质自然引起了研究者的兴趣, 其中一个研究方法便是 Efron 提出的自助法. 传统的自助法主要应用于独立序列样本的重复抽样, 但由于忽视了时间序列相关性, 并不适合长记忆特点的时间序列重复抽样. Pilar<sup>[2]</sup> 分析了一些长记忆检验方法的缺点, 用 Monte Carlo 进行模拟, 比较了不同自助法的检验水平与权势, 结果表明, 滑动分块自助法 (MBB) 置信区间表现更好. 本文引进适合非独立同分布重复抽样的 MBB 方法, 对中国股市进行长记忆的 GPH 检验.

## 1 GPH 检验方法

分整自回归移动平均 (ARFIMA) 模型过程可表示为

$$(B)(1-B)^d X_t = (B)\varepsilon_t, \quad (1)$$

在式 (1) 中,  $\varepsilon_t \sim iid(0, \sigma^2)$ ,  $B$  为滞后算子,  $\Gamma(\cdot)$  为 Gamma 函数,  $(B)$  为分数差分算子, 有

$$(B) = 1 - \phi_1 B - \dots - \phi_p B^p, \quad (B) = 1 - \beta_1 B - \dots - \beta_q B^q,$$

$$(1-B)^d = \frac{\sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k}{(k+1)^{-(d)}},$$

当  $d=0$ , ARFIMA 模型转化为自回归-滑动平均 (ARMA) 模型; 当  $-0.5 < d < 0.5$  时, ARFIMA 过程是稳定的; 当  $d < 0.5$  时, 零均值的 ARFIMA 模型是协方差平稳的; 当  $0 < d < 0.5$  时, 过程是长记忆的; 当  $-0.5 < d < 0$  时, 过程是中程记忆; 当  $d > -1$  时, 过程是可逆的. 因此, 时间序列是否具有长期记忆决定于  $d$  的估计值.

**收稿日期:** 2008-02-29

**通信作者:** 胡日东 (1964-), 男, 教授, 博士生导师, 主要从事金融工程与金融管理的研究. E-mail: j\_rdh@hqu.edu.cn.

**基金项目:** 国家软科学研究计划项目 (2008GXSD130); 国家社会科学基金资助项目 (08BJL019)

对于长记忆模型 ARFIMA( $p, d, q$ ) 的参数  $d$ , Geweke 等<sup>[3]</sup> 采用谱方法(GPH 法) 对其进行估计. 设  $x_t$  在频率处的谱密度和周谱分别为  $g_u(\cdot)$  和  $I(\cdot)$ , 则

$$\ln I(\cdot) = \ln g_u(0) - d \ln[4 \sin^2(\frac{\cdot}{2})] + \ln \frac{I(\cdot)}{f_x(\cdot)}. \quad (2)$$

式(2)中,  $I(\cdot) = \frac{1}{2T} \left| \sum_{i=1}^T \exp(it_j) (x_t - \bar{x}) \right|^2$ ,  $f_x(\cdot)$  为  $x_t$  的谱密度函数. 这意味着参数  $d$  可以通过一个简单的回归方程来估计得到, 即  $\ln I(\cdot) = d_0 - d \ln[4 \sin^2(\frac{\cdot}{2})] + e_j$ . 其中,  $e_j = \ln \frac{I(\cdot)}{f_x(\cdot)}$ . 在调和频率处,  $\cdot$  是渐进独立同分布的,  $n$  为  $T$  的递增函数, 但远远小于  $T$ . Geweke 等<sup>[3]</sup> 建议, 取  $n = T$ ,  $0 < \cdot < 1$ . 若过程存在长记忆, 则  $d$  的最小二乘估计值应显著地不为零.

## 2 滑动分块自助法

滑动分块自助法, 是一种适合具有非独立分布的时间序列的非参数再抽样方法, 能有效地拟合参数的抽样分布<sup>[4-5]</sup>. 给定时间序列  $X = \{x_1, x_2, \dots, x_T\}$ . (1) 构造滑动分块  $\{B_1, B_2, \dots, B_{T-b+1}\}$ , 其中,  $B_i = \{x_i, x_{i+1}, \dots, x_{T-b+1}\}$ . (2) 从  $\{B_1, B_2, \dots, B_{T-b+1}\}$  中随机放回地抽出  $k = \lfloor T/b \rfloor$  个滑动分块  $\{B_{i,1}, B_{i,2}, \dots, B_{i,k}\}$ , 并首尾连接成一个新的样本  $X^{*i}$ ,  $i$  表示第  $i$  次抽样. 这样就可以得到一系列的自助法样本, 利用这些样本去估计参数  $d$ , 可得到  $d$  的自助法置信区间. 具体实现有以下 7 点步骤. (1) 利用原始数据进行 GPH 检验, 得到参数  $d$  的估计值  $\hat{d}$ . (2) 拟合 ARFIMA( $p, \hat{d}, q$ ) 模型,  $p, q$  由赤池信息量(AIC) 准则决定. 对估计模型后得到的残差进行中心化. (3) 将中心化后的残差进行滑动分块自助仿真, 得到  $B$  个自助法样本, 记为  $x^b$ . (4) 根据估计 ARFIMA 得到系数, 以及得到自助法样本, 可以得到  $B$  个原始数据的自助法样本, 记为  $x^b$ . (5) 对于每一个自助法样本  $x^b$ , 重新估计  $d$ , 就可以得到统计值  $\hat{d}^b$ . (6) 对  $B$  个统计值  $\hat{d}^b$  按升序排列, 找出 2.5% 与 97.5% 的分位点, 就可以构造 5% 水平的置信区间. (7) 根据  $\hat{d}$  是否在这个区间内, 判断是否拒绝原假设. 为了精确起见, 所有的模拟结果均为使用  $B = 1\,000$  个自助法样本及有关统计量计算而得.

需要说明的是, 滑动块长度  $b$  的选择是一个影响再抽样好坏的关键因素. 理论上随着原序列长度  $T$  的增大,  $b$  也需增大. 给定样本量  $T$ , 如果  $b$  太小, 则再抽样样本不能模拟原序列的行为, 破坏原序列结构; 如果  $b$  太大, 再抽样样本随机性被破坏. 因此,  $b$  的选择必须平衡二者. 有许多复杂的方法选择滑动块长度  $b$ , 如可采用文[6]提出的公式, 即  $b = CT^{1/3}$ . 其中,  $T$  为样本容量,  $C$  为某个常数. Patton 根据此方法编写了估计最优块长参数  $b$  的 Matlab 程序, 因此,  $b$  的选择采用了这个程序.

## 3 实证检验

众所周知, 如果一个资本市场是有效的, 则市场收益率  $r_t$  自身几乎不具序列相关性. 然而, Taylor 发现当收益率取绝对值后, 序列  $|r_t|$  具有显著的自相关性, 且对于很大的滞后阶数, 序列仍有正的自相关系数. Ding 等<sup>[7]</sup> 通过对美国股票市场标准普尔 500 股指每日收益率的分析发现, 尽管股指收益率序列的自相关性很弱, 但对序列  $|r_t|^d$ ,  $d > 0$  却有很强的自相关性; 特别是当  $d = 1$  时,  $|r_t|^d$  的自相关系数当滞后阶数大于 2 500 时仍为正, 即表现出很强的长记忆性. 借鉴国内外的研究成果, 选取沪、深两市的日指数收益率序列及其变形序列(绝对值收益率)、收益率平方序列进行长记忆检验.

### 3.1 数据来源与描述统计

采用 1997 年 1 月 1 日至 2006 年 12 月 29 日的深圳成分指数与上证综合指数的收盘价作为样本数据, 共 2 401 个样本点(数据来源于 <http://stock.sohu.com>). 设  $t$  时刻的收盘价记为  $p_t$ , 定义日收益为对数收益  $r_t = \ln p_t - \ln p_{t-1}$ . 记深、沪两市的日指数收益率为  $r_{sz}$  和  $r_{sh}$ , 绝对值收益率记为  $|r_{sz}|$  和  $|r_{sh}|$ , 收益率的平方记为  $r_{sz}^2$  和  $r_{sh}^2$ .

表 1 为沪深股市的日指数收益率、收益率绝对值、收益率平方的描述性统计. 从表 1 可以看出, 两市的日指数收益率、收益率绝对值、收益率平方的偏度系数均大于 0, 表现出右偏特征; 而峰度系数均远大于 3, 表明它们的分布有比正态分布更厚的尾部; 同时, J-B 统计量在 1% 的显著性水平下, 拒绝股市收

益率序列及其变形序列服从正态分布的假定.

表 1 各种收益率的描述性统计

Tab. 1 A variety of return rate descriptive statistics

参数	均值	标准差	偏度系数	峰度系数	J-B 统计值	P 值
$r_{sz}$	0	0.016	0.071	7.969	2 471.1	0
$r_{sh}$	0	0.015	0.055	8.804	3 370.4	0
$ r_{sz} $	0.011	0.012	2.622	13.489	13 775.2	0
$ r_{sh} $	0.010	0.011	2.775	15.358	18 353.8	0
$r_{sz}^2$	0	0.001	7.009	67.034	429 723.4	0
$r_{sh}^2$	0	0.001	7.869	82.253	652 348.7	0

图 1 为两市各种收益率的 200 阶滞后的自相关图. 从图 1(a),(d) 可以看到, 两市的收益率 200 阶滞后的自相关函数值( ) 时正时负, 很难说它的衰减过程是一个缓慢的双曲线衰减过程, 明显不太符合长记忆的特征. 图 1(b),(c) 中, 沪市序列  $|r_{sh}|$  与  $r_{sh}^2$  的 200 阶滞后的自相关函数图具有相似的特点, 即自相关函数缓慢递减. 图 1(e),(f) 中, 同样观察到深市的序列  $|r_{sz}|$  与  $r_{sz}^2$  的 200 阶滞后的自相关函数也有相似的特点. 这表明, 沪、深两市两个序列可能是长记忆过程. 当然, 通过图像的观察只是一种经验的判断, 需要进一步采用 GPH 检验进行分析.

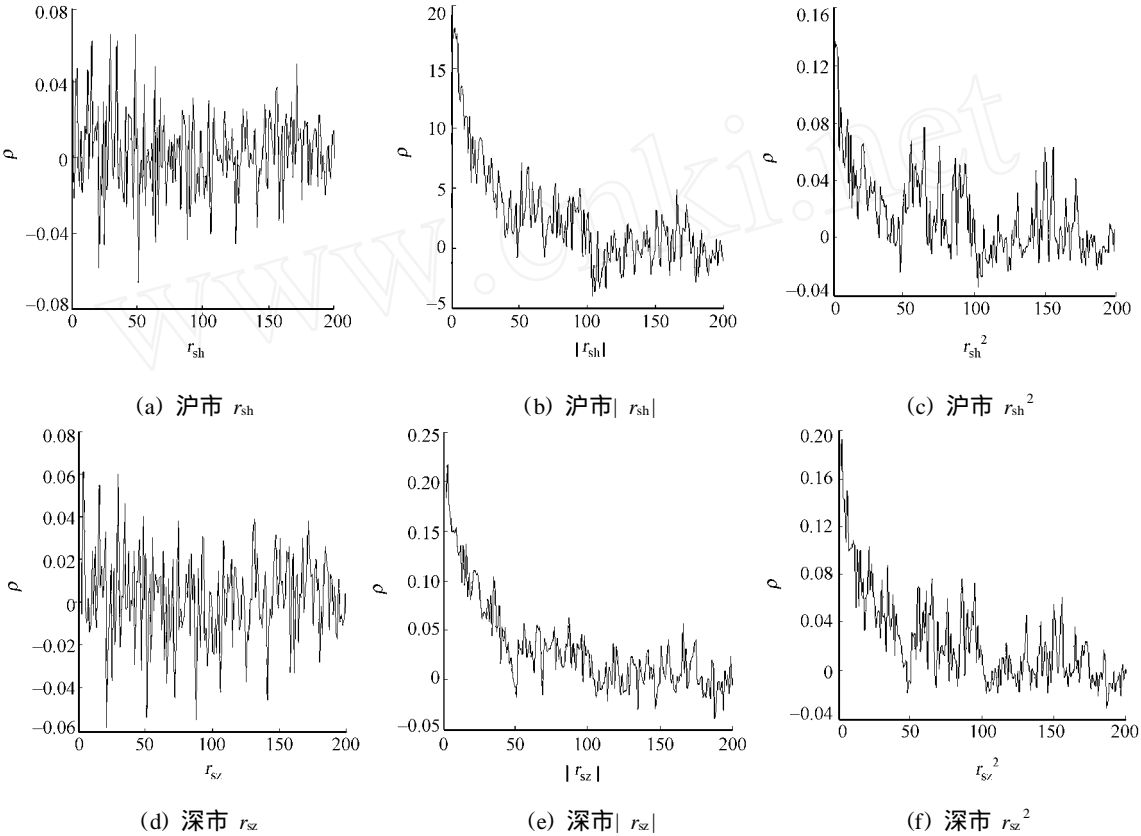


图 1 各种收益率 200 阶滞后的自相关系数

Fig. 1 A variety of return rate autocorrelation coefficient at lag 200

3.2 单位根检验

采用 ADF(Augumented Dickey-Fuller) 单位根检验和 KPSS(Kwiatkowski, Phillips, Schmidt and Shin Test) 单位根检验, 对以上 6 个序列进行平稳性检验. 前者的原假设为序列是一个单位根过程, 而后的原假设序列为平稳序列. 两种单位根检验的结果, 如表 2 所示 (只给出沪市的检验结果). 表 2 中, 单位根检验的最优滞后期由贝叶斯信息准则(BIC) 决定. 由表 2 可知, 无论是带常数项的, 还是同时无常数项和趋势项的情形, 在 5 % 显著性水平下, ADF 检验拒绝原假设, 沪市 3 个收益率序列是平稳序列. KPSS 检验在 5 % 显著性水平下也接受沪市收益率序列是平稳序列的原假设, 但对于序列  $|r_{sh}|$  与序列  $r_{sh}^2$ , 检验结果拒绝了平稳序列的原假设. 综合来看, 沪市的序列  $r_{sh}$  是一个  $I(0)$  过程, 但是序列  $|r_{sh}|$  与序

列  $r_{sh}^2$  可能不是单纯的  $I(0)$  或者  $I(1)$  过程,而可能是一个分数差分过程.

表 2 沪市 3 种收益率单位根检验结果

Tab.2 The results of Shanghai stock market 's three kinds return rates unit roots test				
序列	统计值与临界值	ADF 单位根检验		KPSS 单位根检验
		常数项	无常数与趋势项	无常数与趋势项
$r_{sh}$	统计值 5 %显著性水平临界值	- 48.80	- 48.77	0.216
		- 2.86	- 1.94	0.463
$ r_{sh} $	统计值 5 %显著性水平临界值	- 15.04	- 3.74	0.651
		- 2.86	- 1.94	0.463
$r_{sh}^2$	统计值 5 %显著性水平临界值	- 18.54	- 8.39	0.811
		- 2.86	- 1.94	0.463

为验证此猜想,考虑对序列进行分数差分,取  $d=0.1,0.3,0.5,0.7$ . 令  $Y_t$  为分数差分后得到的序列,即  $Y_t=(1-L)^dX_t$ . 其中,  $X_t$  是沪、深两市的收益率序列及其变形序列. 然后,对  $Y_t$  进行单位根检验. 结果如表 3 所示. 因篇幅限制,只给出两市绝对值收益率序列的检验结果.

对沪市序列  $|r_{sh}|$  分数差分后的单位根检验后,可以发现,  $d=0.1$  差分后得到的序列仍为非平稳序列,而  $d=0.5$  差分后得到了平稳的序列. 这表明  $d=0.7$  已过度差分,若对序列  $|r_{sh}|$  进行普通的差分,将造成过度差分,损失序列中的有用信息. 这进一步支持沪市  $|r_{sh}|$  序列是一个分数差分过程. 预计  $d$  值处于 0.3 与 0.5 之间,但具体的  $d$  值还需要根据 GPH 检验法估计得到. 同样,从表 3 还可以发现,对深市  $|r_{sz}|$  序列进行  $d=0.1$  分数差分为不足差分,而  $d=0.7$  分数差分已经是过度差分. 这表明深市  $|r_{sz}|$  序列也是一个分数差分过程.

表 3 收益率绝对值分数差分序列的单位根检验

Tab.3 The unit roots test of fractional differencing absolute return rate series								
$Y_t=(1-L)^d r_t $	沪市				深市			
	$d=0.1$	$d=0.3$	$d=0.5$	$d=0.7$	$d=0.1$	$d=0.3$	$d=0.5$	$d=0.7$
KPSS 检验统计值	0.376	0.352	0.042	0.031	0.460	0.401	0.039	0.012
10 %水平临界值	0.347	0.347	0.347	0.347	0.347	0.347	0.347	0.347
检验结论	非平稳	非平稳	平稳	平稳	非平稳	非平稳	平稳	平稳

3.3 滑动分块自助法仿真临界值的确定

根据谱回归方程(2)估计分数差分参数时,需要先选取回归方程中包含的低频周期图分量数  $n$ . Kunsch 的研究表明,在  $n=T^{0.5}$  ( $T$  为样本容量)附近时估计结果更准确. 结合其他文献中关于选取低频周期图分量数的讨论,并考虑到认识估计分数差分参数对回归样本规模敏感性的需要,对  $n=T^{0.5}, n=T^{0.525}, n=T^{0.55}, n=T^{0.575}$  和  $n=T^{0.6}$  5 种情形分别估计了分数差分参数,并根据 MBB 的实现步骤得到分数差分参数  $d$  的 5 %水平的 MBB 仿真临界值. 表 4 为深市  $r_{sz}, |r_{sz}|, r_{sz}^2$  序列的长记忆自助法检验.

从表 4 可以看出,对于不同低频周期图分量数,深市收益率序列参数  $d$  的估计值变化不大,但是其对应的  $T$  值在 5 %水平上并不显著;从构造的 5 %水平自助法仿真临界值上看,各  $d$  值也处于接受域内. 这两点都表明,深市的收益率序列不具有长记忆的特征. 这也支持了前面的分析,同时也表明, GPH 检验与滑动分块自助法的 GPH 检验的结论是一致的.

表 4 深市各序列长记忆自助法检验

Tab.4 Long-term memory bootstrap test of a variety of Shenzhen stock return rate										
$n$	$r_{sz}$			$ r_{sz} $			$r_{sz}^2$			
	$d$	$T$ 值	5 %临界值	$d$	$T$ 值	5 %临界值	$d$	$T$ 值	5 %临界值	
$n=T^{0.500}$	0.117	0.918	- 0.154, 0.182	0.358	4.24	- 0.144, 0.350	0.347	3.00	0.128, 0.324	
$n=T^{0.525}$	0.130	0.855	- 0.131, 0.173	0.326	4.34	- 0.099, 0.316	0.315	3.25	0.018, 0.308	
$n=T^{0.550}$	0.121	1.004	- 0.111, 0.180	0.331	4.68	- 0.043, 0.323	0.369	3.94	0.046, 0.357	
$n=T^{0.575}$	0.167	0.897	- 0.103, 0.185	0.388	4.63	0.015, 0.378	0.332	3.61	0.023, 0.312	
$n=T^{0.600}$	0.154	1.689	- 0.080, 0.167	0.359	5.33	0.076, 0.346	0.364	4.66	0.052, 0.346	

同样也发现,沪市的收益率序列也不是一个长记忆过程. 从表 4 还可以看出,深市  $|r_{sz}|, r_{sz}^2$  序列是一个

长记忆过程,分数差分参数大致在 0.35 左右.沪市这两种序列的  $d$  值大致在 0.3 左右.这表明深市的两种序列的长记忆性略比沪市更长.值得注意的是,虽然估计差分参数对谱回归方程中包含的低频周期图分量数有一定的敏感性,但都得到了比较精准的自助法仿真临界值.

## 4 结束语

滑动分块自助法是一种适合自相关时间序列的再抽样方法,与传统自助法仿真相比,它能有效地拟合同统计量的抽样分布.在对中国股市长记忆进行 GPH 检验时,采用一种在重复抽样时保持误差自身结构的自助仿真方法——滑动分块自助法,对分整参数  $d$  构造精准的自助法置信区间.实证结果显示,沪深股市的日指数收益率不具有长记忆特征,而收益率绝对值与收益率平方序列具有长记忆特征.沪市收益率绝对值和收益率平方序列的分数差分值大约为 0.3,深市大约为 0.35.因此,深市的长记忆强于沪市.这基本符合两市的实际情况.

### 参考文献:

- [1] 陈梦根.中国股市长期记忆的实证研究[J].经济研究,2003(3):70-78.
- [2] PILAR G C. Tests of long memory: A bootstrap approach[J]. Computational Economics,2005(25):103-113.
- [3] GEWEKE J, PORTER-HUDAK S. The estimation and application of long memory time series models[J]. Journal of Time Series Analysis,1983(4):221-238.
- [4] 张术林,魏正红.金融资产收益非对称性分析[J].深圳大学学报:社会科学版,2007(1):81-84.
- [5] RICHARD M V. The moving blocks bootstrap versus parametric time series models[J]. Water Resource Research,1996,32(6):1875-1882.
- [6] POLITIS D N, WHITE H. Automatic block-length selection for the dependent bootstrap[J]. Econometrics Reviews,2004,23(1):53-70.
- [7] ZHUAN Xin-ding, CLIVE W J G. Modelling volatility persistence of speculative returns: A new approach[J]. Journal of Econometrics,1996(73):185-215.

## The Test of Long Memory of China's Stock Market Based on Moving-Block Bootstrap Simulation

SU Zhi-fang, HU Ri-dong, CHEN Jia-gan

(College of Commerce, Huaqiao University, Quanzhou 362021, China)

**Abstract:** This paper proposes a test of GPH (Geweke, Porter-Hudak) based on the moving-block bootstrap approach and tests the long-term memory of daily return rate sequence both in Shanghai stock market and Shenzhen stock market, which shows that daily return rate sequence is an  $I(0)$  process and not a long-term memory. However, the absolute return rate and squared return rate sequence are fractional differencing process. The fractional differencing parameter of absolute return rate and squared return rate sequence in Shanghai stock market is about 0.35, and the fractional differencing parameter of Shenzhen's is about 0.30, i.e. the long-term memory of Shenzhen stock market is stranger than Shanghai's, and it also indicates that Shanghai stock market is more efficient than Shenzhen's.

**Keywords:** long-term memory; auto-regressive fractiona linteegrated moving average; moving-block bootstrap approach; GPH test; China stock market

(责任编辑:钱筠 英文审校:司福成)