

# A股市场长记忆性特征研究

王谦<sup>1</sup>, 刘春<sup>1</sup>(副教授), 管河山<sup>1</sup>(副教授), 罗智超<sup>2</sup>(高级工程师)

(1.南华大学经济管理学院, 湖南衡阳 421001; 2.厦门大学王亚南经济研究院, 厦门 361005)

**【摘要】**本文采集了我国A股市场所有上市公司的价格数据和收益率数据,分别采用R/S、MR/S和V/S检验方法进行实证研究,发现:①数据平稳性特征会对长记忆性检验结果造成直接影响,而股票价格和收益率数据不都是平稳的,因此,应该事先对数据进行平稳性检验;②不同股票样本、不同检验方法得到的检验结果有较大差异,根据市场指数或部分股票样本探究股市长记忆性有一定的片面性;③同一时间段,数据采样间隔(日、周)会影响到数据总量,进而造成长记忆性检验结果的差异。对A股市场的股票总体分析表明,无法给出市场长记忆特征的最终结论。因此,相对股市整体而言,探究个股长记忆性特征更具有研究价值。

**【关键词】**A股市场;长记忆性;检验方法

## 一、引言

英国水利学家Hurst(1951)对水文数据特征的分析开启了时间序列记忆性特点研究的大门。随后,时间序列长记忆性特征分析被广泛地运用到金融领域,这为准确把握金融市场的运行规律,建立符合其特征的精确预测模型提供了现实可能性。长记忆性作为金融市场价格波动的重要研究内容之一,它侧重探究价格序列本身在较远时间间隔的情况下是否仍具有持续依赖关系。而时间序列长记忆性特征的存在,使得利用时间序列本身的历史信息来预测其未来变化趋势成为可能,它作为建立预测模型的关键因素之一,对金融资产价格记忆性及其本质特征的研究不仅可以为上层政策制定和宏观调控提供决策依据,还能为组织和个人投资者提供一定现实参考。

近年来,关于我国金融市场长记忆性特征的研究颇多。比如,何兴强(2006)和顾荣宝(2008)对我国沪深股市进行V/S检验,认为我国股票市场呈现反持续性特征而不是长记忆性特征;许友传(2010)发现R/S检验更倾向于接受长记忆性,并得出需谨慎解释采用R/S检验方法和修正的R/S检验方法所得出实证结果的结论;李志生(2011)采用序列相关检验和R/S检验法分析得出我国股票市场有显著的长记忆性并在近年来逐步增强的趋势;苑莹(2012)运用R/S法、MR/S法、KPSS检验及Granger因果检验方法发现我国股市收益率和成交量均具长记忆性特征;肖炜麟(2013)研究了长记忆建模影响问题,发现不同模型的定价结果证明了金融资产具有长期记忆性。庞淑娟(2011)、陈浪南(2013)等对我国股市高频数据的长记忆性特征进行了研讨。楼晓东(2013)提出V/S方法在识别长记忆性时最贴近实际,并通过计算出Hurst指数证实

了国际黄金现货(周线/月线)存在着显著的长记忆性,表明黄金市场不是弱势有效的。

上述研究虽然拓展了长记忆性的研究,然而,学者们在对我国股市长记忆性的研究过程中,其实证分析中都采用抽样样本进行分析,研究结论具有不一致性。部分学者认为金融股市不具有长记忆性,部分学者认为金融股市具有长记忆性,部分学者针对不同地区的金融股市进行分析,认为发达地区股市一般不具有长记忆性,而发展中国家股市具有长记忆性。可见,在不同股市环境、不同数据采集时段、不同检验方法对检验结果有一定的影响,学者们对金融市场的长记忆性判定各执一词,其研究结论的可靠性仍需要深入探讨。为此,本文采用长记忆性检验中常用的三种方法(重标极差检验法、修正的重标极差检验法和重标方差法),并采集我国A股市场所有上市公司的数据进行分析,全面探究A股市场的长记忆性特征。

## 二、长记忆性定义及检验方法

### (一)长记忆性定义

长记忆性(long-term memory)又称长程相关性或持久性,它描述了时间序列在较远间隔处的持续依赖关系,表明时间序列具有非周期的循环特征,如果其动态相依结构表现出某种长记忆特征或长时相依关系,则时间序列的历史信息将有助于预测其未来变化趋势。国内外对长记忆的定义有很多种,本文选取两种较常见的定义方式,如下所示:

第一种是根据自相关函数的定义,当平稳过程 $X_t(t=1, 2, \dots, T)$ 的自相关函数 $\alpha(k)$ 满足:

$$\alpha(k) \rightarrow C_\alpha k^{2d-1} \quad (1)$$

其中 $k \rightarrow \infty, C_\alpha > 0, d < 1/2$ 则认为平稳时间序列 $\{X_t\}$

具有长记忆性或长相依性。

第二种是谱密度函数定义,如果存在一个实数  $d$  和一个有界常数  $C$ ,使得对于一个平稳序列  $\{X_t\}$ ,它的谱密度  $f\{\rho_j\}$  满足:

$$f\{\rho_j\} \sim C \times \rho_j^{-2d} \quad \rho \rightarrow 0 \quad (2)$$

其中  $d \in (0, 0.5)$ ;  $\rho_j = 2\pi \times j / N$ ,  $j = 1, 2, \dots, m$ ;  $N$  为样本总数,  $m$  为带宽,  $m = g(T) \leq T$ , 则认为时间序列  $\{X_t\}$  具有长记忆性或长相依性。

值得注意的是:进行长记忆性分析的前提条件是时间序列数据满足平稳性的假设,然而现实中的股票价格和收益率等金融时间序列数据有很多是非平稳的,因此,在进行长记忆性分析之前探讨数据的平稳性特征就显得十分必要。平稳性假设主要是为了保证在实证分析中可以随意运用在此假设下各种成熟的检验方法。Mikosch 等(2004)、李锐(2007)发现,如果平稳性假设得不到满足可能会导致错误的检验结果,这使得讨论长记忆性检验方法在非平稳框架下的稳健性成为必要,可见,数据的平稳性特征是检验其是否满足长记忆性特征的重要前提条件。

### (二)长记忆性检验方法

1. 重标极差检验法(R/S)。对时间序列长记忆性的理论研究最早始于20世纪50年代。Hurst(1951)在对水文数据的研究中发现了时间序列具有记忆性的属性,由此提出了重标极差分析的检验方法,并给出了相应的判别式。给定一个时间序列  $\{X_t\}$ ,  $t = 1, 2, \dots, T$ , 将原始时间序列等分成  $A$  段,即  $T = A \times n$ ,  $n$  为子序列长度。再求解回归方程的斜率线  $H$  作为 R/S 的检验量,如下所示:

$$\ln(E(Q_n)) \approx \ln C + H \times \ln(n) \quad (3)$$

$$Q_n = R_n / S_n$$

$$R_n = \max_{1 \leq k \leq n} \sum_{j=1}^k (x_j - \bar{x}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^k (x_j - \bar{x}_n)$$

$$S_n = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2}$$

$$\text{其中 } \bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j$$

根据 R/S 方法求得的  $H$  值,并进行长记忆性判定,判定原则如下:

(1)当  $0 \leq H < 0.5$  时,时间序列呈现出反持久的分形布朗运动的特点。

(2)当  $H = 0.5$  时,标准随机游走序列,不同的时刻点值无关。

(3)当  $0.5 < H < 1$  时,具有长记忆性的时间序列。

(4)当  $H = 1$  时,时间序列为一条直线,未来完全可以预测。

(5)其他,长记忆性判定失效。

理论上来说,R/S方法在分析存在短期记忆性或具有异方差的时间序列时,会更多地倾向于拒绝,不存在长记

忆性的原假设,而这与本方法统计量构造的数学原理有密切关系,因此 R/S 方法在进行长记忆性检验时很容易出现偏倚或错估。还有可能的情况是在对 R/S 统计值进行回归求解时, $H$  值可能出现负值, $H$  值不偏不倚等于 0.5 的情况很难出现,因此对  $H$  值的判断标准可能并不能涵盖所有情况。

2. 修正的重标极差检验法(MR/S)。W. Lo(1991)在其论文《股票价格市场的长记忆性》一文中描述到当序列包含短记忆性、存在异质性、具有非平稳性时,R/S 分析的结果就会大打折扣,因此 W. Lo 也提出了相应的修正方法,即修正的 R/S (MR/S),以此来弥补原有方法判断时间序列长记忆性上的缺陷。具体修正如下:

$$Q_n = R_n / \theta_n(q) \quad (4)$$

$$\theta_n(q) = S_n^2 + 2 \sum_{j=1}^q W_j(q) \times \gamma_j$$

$$W_j(q) = 1 - \frac{j}{q+1} \quad (q < n)$$

$$\gamma_j = \frac{1}{n} \sum_{i=j+1}^n (x_i - \bar{x}_n)(x_{i-j} - \bar{x}_n)$$

其中  $R_n$ 、 $S_n^2$  和  $\bar{x}_n$  的定义与 R/S 中的相似, $\gamma_j$  为样本的自协方差。

由上式可以看出,MR/S 法与 R/S 法虽在分母的计算上有明显的区别,但通过引入调整项  $\theta_n(q)$  可以在一定程度上修正 R/S 过度拒绝时间序列存在长记忆性的原假设,使得结果的输出更为无偏有效。但式中  $q$  值的确定难以统一, $q$  值太小时不利于消除序列短期记忆对检验结果的影响,而  $q$  值太大又会影响该方法的检验效率和性能,尽管 W. Lo 在 1991 年的论文中又提出了  $q$  值的经验计算公式。但不难发现,经验值  $q$  并不是十分可靠的,当子序列长度  $n$  值较小时,可能出现  $q > n$  的情况,造成检验结果的失效。本文计算  $q$  值,借鉴许友传(2010)的方法,得到如下计算公式:

$$q = 4 \times \left( \frac{T}{100} \right)^{\frac{2}{9}} \quad (5)$$

3. 重标方差法(V/S)。由于前文中提到的 MR/S 法在  $q$  值的确定上经验值存在缺陷,而在学术界对于  $q$  值的界定和选取又缺乏统一公认的标准,Giraitis 等人(2003)提出了一种明显异于前两者的新的检验时间序列长记忆性的方法——重标方差法。V/S 法的基本思想是:将总长度为  $M$  的时间序列  $\{X_t\}$  划分为  $A$  个长度为  $N$  的等长子区间,记作  $D_a$ ,  $a = 1, 2, \dots, A$ 。子序列  $D_a$  对应的元素为  $x_{i,a}$ ,  $i = 1, 2, \dots, n$ ,子序列均值为  $\bar{x}_a$ ,子序列标准差为  $S_{n,a}$ 。V/S 法的判别式如下:

$$(V/S)_n = \frac{1}{n \times S_n^2} \left\{ \sum_{k=1}^n \left( \sum_{i=1}^k (x_{i,a} - \bar{x}_a) \right)^2 - \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^k (x_{i,a} - \bar{x}_a)^2 \right\} \quad (6)$$

其中：

$$S_n = \sqrt{\frac{1}{n} \sum_{i=1}^k (x_{i,a} - \bar{x}_a)^2}$$
 为子序列标准差

$$\bar{x}_a = \frac{1}{n} \sum_{i=1}^k x_{i,a}$$
 为子序列均值

对于不同区间划分  $N$  有不同的子序列的  $V/S$  值,而  $V/S$  统计量满足于  $(V/S)_{n=k \times n^{2H}}$  的函数关系,可通过对  $\log(n) \sim \log(V/S)_n$  进行最小二乘估计来求得直线的斜率,而  $V/S$  法的  $H$  值即为回归斜率的一半。一般认为最新的  $V/S$  法在构造机理上的优势有效地避免了  $R/S$  分析的不足,使得该方法在判断时间序列长记忆性上更为稳健,其检验结果也更为有效可靠。但不难发现,  $V/S$  法统计量计算复杂度为  $O(n^2)$ ,高复杂度使得其计算速度和效果不尽如人意,尤其在金融高频数据处理过程中,这种不足会愈加明显和突出,甚至导致实验中断,软件系统在有限的内存下运算容易陷入瘫痪。

### 三、现有研究的评述

1.  $R/S$ 、 $MR/S$  和  $V/S$  方法是长记忆性研究中常用的三种方法,从理论上分析而言: $R/S$  方法通常倾向于拒绝不存在长记忆性的原假设,这与  $R/S$  检验统计量的构造原理有关; $MR/S$  方法通过对  $R/S$  检验统计量分母的长期方差引入,从而降低了其对短记忆和异方差的敏感度,提高了模型的适应性,但  $MR/S$  检验结果依赖于模型的滞后阶数选择,  $V/S$  方法对时间序列短期相关性的敏感度却没有明显的优势,且计算复杂度高。因此,以上检验方法各有利弊,综合采用多种检验方法来进行综合判定,避免采用某种单一的方法进行武断判定,这已经成为众多学者的共识。

2. 现有文献的实证研究大多采用了特定金融市场的综合指数或某些典型上市公司的数据来进行分析,缺乏对 A 股市场所有上市公司的全面分析,因此,典型样本分析的结论是否能如实反映市场总体的特征仍值得探究。

3. 长记忆性检验过程中,对数据平稳性检验的重视程度不够,甚至部分学者在实证过程中未曾对数据的平稳性进行分析,这会直接影响到长记忆性检验结果的合理性。

4. 部分学者研究发现,数据采样的具体时间段对长记忆性检验结果影响较大,但是,对同一采样时段而言,不同的数据采样间隔(日数据、周数据等)是否对长记忆性特征形成影响,仍需进一步探究。

5. 尽管大部分学者都认为单一检验方法的结果存在片面性,然而它们对股市长记忆性判断的结果,其差异到底有多大,是否会对检验结果形成本质性的差异,仍值得深究。

### 四、我国 A 股市场的长记忆性特征实证分析

#### (一)数据说明

采集我国 A 股市场所有上市公司(含市场指数,共

2 597 个数据样本)的收盘价和收益率(考虑现金红利再投资的日个股回报率),采样时段分别为 1991 年 1 月 1 日~2014 年 4 月 30 日(全时段)、2006 年 1 月 1 日~2008 年 12 月 28 日(牛市)和 2009.1.1-2011.12.31(熊市),数据来源于 CSMAR 数据库,借助 SAS 软件处理数据,分别采用  $R/S$ 、 $MR/S$  和  $V/S$  方法进行分析。实证时,子序列长度  $n$  的取值范围定为  $[10, \text{int}(T/2)]$ ,  $T$  为原始时间序列的时间长度;当采用  $MR/S$  法进行分析,需要计算参数  $q$  的取值,为此我们对子序列长度  $n$  的最小取值进行了调整,调整为  $[n_0, \text{int}(T/2)]$ ,  $n_0 = \max\{10, q\}$ 。

#### (二)平稳性检验分析

根据长记忆性的定义,需要假定数据本身具有平稳性。采用 ADF 方法检验其平稳性,结果如表 1 所示:

数据平稳性特征	全时段价格数据				价格数据		收益率数据	
	日	周	日	周	牛市	熊市	牛市	熊市
平稳的	1 797	823	2 597	2 580	670	165	2 349	1 361
非平稳的	800	1 757	0	0	1 927	2 432	248	1 236

由表 1 可知,对价格数据而言,大部分数据都是非平稳的,而对收益率数据而言,大部分数据都是平稳的。根据收益率的计算原理可知,收益率数据近似于价格数据的对数差分,而差分是消减非平稳性特征的重要途径,因此,收益率数据的平稳性样本更多。

#### (三)长记忆性分析

根据表 1 的分析结果,进一步提取平稳数据分析。对三种采样时段(全时段数据、牛市和熊市)而言,只要公司上市时间较早,则其样本的数据长度都超过三位数字,分别采用  $R/S$ 、 $MR/S$  和  $V/S$  法展开实证,分析结果见表 2:

方法	检验结果	全时段数据				牛市		熊市	
		日价格	日收益率	周价格	周收益率	日价格	日收益率	日价格	日收益率
$R/S$	长记忆	0.58	0.834	0.502	0.781	0.085	0.616	0.564	0.769
	反持久	0	0.163	0	0.196	0.001	0.07	0	0.231
	判定失败	0.42	0.002	0.498	0.023	0.913	0.313	0.436	0
$MR/S$	长记忆	0.973	0.471	0.646	0.087	0.143	0.083	1	0.072
	反持久	0.027	0.529	0.293	0.89	0.022	0.605	0	0.928
	判定失败	0	0	0.061	0.023	0.834	0.313	0	0
$V/S$	长记忆	-	-	0.096	0.574	0.016	0.471	0	0.677
	反持久	-	-	0	0.402	0.001	0.215	0.024	0.323
	判定失败	-	-	0.904	0.024	0.982	0.314	0.976	0

需注意表2中的数据:(1)V/S法的计算复杂度较高,日采集方式的数据量太大,因此未采用V/S法分析;(2)部分个股的上市时间较短,数据总量不够,出现判定失误的情况,将其并为“判定失败”一栏。

由表2分析可知:

1. 对价格和收益率两种数据类型进行分析,价格数据中反持久特征的样本比例明显偏低,而收益率数据中反持久特征的样本比例明显偏高;收益率数据可以近似看作价格数据的对数差分,其随机波动性更强,造成了收益率数据呈现出更为明显的反持久分形布朗运动的特点。

2. 对全时段的日数据和周数据两种数据采集方式进行分析,三种方法的判定结果差异比较明显,这意味着,对同一采样时段而言,不同的数据采集方式,得到的数据长度不同(样本量差异),这将对长记忆性检验结果的产生显著影响,因此,考虑某个时段上市公司的长记忆性特征,应根据实际问题采取相应的数据采集方式。

3. 对长记忆性判定进行分析,R/S、MR/S和V/S法得到结果差异较大:R/S法的结果更倾向于接受长记忆性,特别是对收益率数据而言,这一特点更为明显;相对R/S方法而言,MR/S法检验得到的结果中,更多个股倾向于反持久的分形布朗运动的特点,特别是对收益率数据而言,呈现出明显的反持久的分形布朗运动的特点;V/S法判定失误的概率明显较大,判定结果中一半接受长记忆性,另一半则倾向于反持久的分形布朗运动的特点。综合以上分析,三种方法的检验结果有显著的差异,其结果需要进一步确定。

4. 对两种市场行情(全时段、牛市、熊市)进行分析,排除检验失效的样本,R/S和V/S法对不同市场行情趋势下(牛市和熊市)的长记忆性检验结果的差异较小;MR/S方法对牛市和熊市的两种市场行情趋势的长记忆性检验结果的差异非常明显,该方法在日价格中,样本都表现出较强的长记忆性,然而在日收益率数据下,该方法却更明显地倾向于接受反持久性特征,这与第三个结论是一致的。

## 五、结论

本文借助R/S、MR/S和V/S三种判定方法展开综合分析,并采集我国A股市场中所有个股的历史数据进行实证,综合以上研究结果发现:

1. 对A股市场的个股而言,不管其价格数据,还是收益率数据,都不可能全部是平稳的,因此,对我国A股市长记忆性的研究,应该首先对数据的平稳性进行分析。不可否认,平稳性分析是长记忆性检验的必要预处理过程。

2. 不同的检验方法(R/S、MR/S和V/S)、不同个股得到的长记忆性判定结果有明显的差异,因此,很难给出我国A股市场是否长记忆性特征的综合结论。已有研究中,仅

采用A股市场综合指数或部分典型样本分析并得出股市长记忆性判定的结论,其观点存在一定的片面性。

3. 不同时段(全时段、牛市、熊市)的数据,其长记忆性存在明显差异;然而,对于同一时间段,数据采集间隔(日数据、周数据)的不同也会导致长记忆性检验结果的差异显著,这正是本文的研究结论之一。

股市长记忆分析结果受到数据平稳性特征、数据采集间隔、长记忆检验方法及样本选取的影响。本文通过对A股市场的股票总体分析(而非抽样样本),实证结果无法给出市场长记忆特征的最终结论,因此,相对股市整体而言,探究个股长记忆性特征更具有研究价值。

## 主要参考文献

Hurst H. E., Long-term storage capacity of reservoirs [J]. Trans. Amer. Soc. Civil Eng, 1951(16).

何兴强,李仲飞.上证股市收益的长期记忆:基于V/S的经验分析[J].系统工程理论与实践,2006(12).

顾荣宝,陈霁霞.基于分形V/S技术的沪深股市长记忆性研究[J].安徽大学学报(自然科学版),2008(3).

许友传.金融时间序列R/S长记忆性检验的有效性[J].中国管理科学,2010(8).

李志生,刘正捷.资产收益的短记忆性与长记忆性:我国股票市场效率的动态分析[J].江西财经大学学报,2011(1).

宛莹,杜乐鹿,庄新田.股市收益率与交易量长记忆性实证研究[J].东北大学学报,2012(7).

肖炜麟,张卫国,徐维东.双分式布朗运动下股本权证定价[J].系统工程学报,2013(3).

庞淑娟,刘向丽,汪寿阳.中国期货市场高频波动率的长记忆性[J].系统工程理论与实践,2011(6).

陈浪南,杨科.中国股市高频波动率的特征、预测模型以及预测精度的比较[J].系统工程理论与实践,2013(2).

楼晓东,张良.基于分形理论的国际金价波动长记忆性识别及预测研究[J].上海金融,2013(6).

Daniel O. Cajueiro, Benjamin M. Tabak. The rescaled variance statistic and the determination of the Hurst exponent[J]. Mathematics and Computers in Simulation, 2005(70).

刘春,王谦,管河山.系统协调度模型的稳健性研究[J].南华大学学报(社会科学版),2014(5).

刘湘云,卞悦.我国农产品期货的长记忆研究:基于GPH与修正R/S实证检验[J].南京财经大学学报,2011(3).

【基金项目】教育部人文社科青年项目“海量金融时间序列数据平稳性检验方法研究”(编号:13YJCZH044);湖南省科研创新立项课题“大数据时代金融时间序列长记忆性检验方法研究”(编号:CX2014B397)