

中国股市波动性与成交量共同的长期记忆性研究

张庆翠, 王春峰

(天津大学系统工程研究所, 天津 300072)

摘要: 研究了上证30指数和深圳成分指数所含个股的成交量与收益波动性的长期相关关系。文章应用多变量频域两步半参数估计方法, 辅助数据加窗技术对非平稳时间序列的长期记忆性进行了一致有效的估计。研究结果表明中国股票市场的成交量和波动性均具有显著的长期记忆性, 并且对于大多数股票而言, 成交量序列与波动性序列具有相同的长期记忆性。表明同时驱动成交量和收益波动性的潜在信息流本身具有长期记忆特征。

关键词: 长期记忆性; 非平稳随机过程; 谱分析; 加窗周期图

中图分类号: F830.91 **文献标识码:** A **文章编号:** 1007-9807(2005)02-0038-08

0 引言

时间序列具有长期记忆性是指相距较远的时间间隔具有显著的序列相关性。近年来, 对金融市场波动性的研究一直是热点课题, 而越来越多的实证研究结果表明股票市场的波动性过程具有长期记忆和均值回复的特点。国外大量学者作了此方面的工作^[1~3], 国内对股票市场波动性过程的长期记忆性研究相对比较匮乏。李汉东、张世英从理论方法角度介绍了分整自回归条件异方差(FIGARCH)模型^[4]; 王春峰、张庆翠从实证角度对中国股票市场的研究表明, 中国股市和国外股市一样, 波动性过程具有显著的长期记忆性^[5]。

然而, 直至今日, 理论界对于波动性过程长期记忆性的形成机制仍没有达成共识。一种解释将其归因于信息到达过程的不均衡; Andersen 和 Bollerslev 则认为波动性过程的长期记忆性构成了收益生成过程的内在特征; 而另一种解释则把波动性过程的长期记忆性与成交量过程联系起来, 可是成交量和波动性谁因谁果? 还是两者均由一共同因素驱动?

国外大量学者对量价关系的研究结论, 大都支持成交量与价格波动正相关, 大的成交量伴随着大的价格波动。支持这一结论的具有说服力的假说为混合分布假说(the Mixture Distribution Hypothesis) 简计为 MDH^[6]。该假说认为: 资产的回报和成交量是由一个潜在的、不可观测的信息过程(或信息流)共同驱动的, 当非预期的利好消息到达市场时会引起价格上涨; 反之, 当非预期的利空消息到达时会引起价格下跌。无论是价格上涨还是价格下跌, 均将伴随着高于平均水平的交易行为, 直至这一消息被充分吸收, 市场调整到一个新的均衡为止。因此, 成交量和资产回报的绝对值正相关。尽管最初的实证研究结论大多接受这一假说, 但 Lamoureux 和 Lastrapes^[7]的研究表明相同的、潜在的信息到达过程并不能描述成交量和波动性过程的短期相关关系, 相反, 他描述的是一种长期相关关系。即在信息到达市场初期, 成交量和波动性过程没有必要一定做出相同的反应。

然而研究量价长期相关关系, 进而揭示价格波动持续性根本原因的相关文献并不多见。Bollerslev 和 Jubinski^[8]分析了混合分布假说模型, 发现潜在的

收稿日期: 2002-11-25; 修订日期: 2005-01-18.

基金项目: 国家杰出青年基金资助项目(70225002); 教育部高等学校优秀青年教师教学科研奖励计划基金资助项目; 教育部跨世纪优秀人才计划基金资助项目.

作者简介: 张庆翠(1975—), 女, 河北人, 博士.

信息到达过程具有长期记忆的特点. 并通过对美国标准普尔 100 复合指数所包含的 100 支股票进行了实证研究, 发现绝大多数股票的成交量和价格波动性序列具有相同的长期记忆性参数.

鉴于成交量序列的非平稳性和具有长期趋势, 使得长期记忆性的研究更加复杂, Bollerslev 和 Jubinski 采取的做法是抽取一线性趋势. 但在 Lobato 和 Velasco^[9~12] 的研究中却发现这样做是有欠妥当的, 因为首先我们不可能确定趋势的确切形式, 再者去除的趋势又可能正好与估计的长期记忆性参数有关. 为了克服上述缺点, Lobato 和 Velasco 采用了频域加窗周期图估计方法研究了美国道琼斯工业平均指数所包含的 30 支股票, 得出了与 Bollerslev 和 Jubinski 相似的结论: 即成交量序列具有长期记忆性; 大多数股票的波动性序列与成交量序列具有相同的长期记忆性. 而国内对这一领域的研究尚未见有相关文献发表.

本文结合 Bollerslev 和 Lobato 的方法研究了中国股票市场的上证 30 指数所包含的 30 支股票和深证成分指数所包含的 40 支股票中成交量和波动性之间的长程相关关系, 填补了国内在这一研究领域的空白.

1 模型定义与参数估计

1.1 平稳过程的长期记忆性定义

协方差平稳的随机过程具有长期记忆性, 且记忆参数为 d , 若其谱密度函数 $f(\cdot)$ (简称 SDF) 满足

$$f(\cdot) \sim G^{-2d}, \quad 0^+ \quad (1)$$

其中: G 为一正的有限常数, $d \in (-0.5, 0.5)$. 参数 d 控制着序列的记忆程度, $(-0.5, 0.5)$ 为随机过程的平稳可逆区间, 即若 $d < 0.5$, 表明随机过程是非平稳的; 若 $d > 0.5$, 表明随机过程是非可逆的.

1.2 非平稳过程的长期记忆性定义

非平稳随机过程 $\{X_t\}$, 若经 s 阶差分后 $({}^s X_t)$ 为一协方差平稳过程, 则称其具有记忆参数为 $d_X = -0.5$ 的长期记忆性. 即对 $U_t^{(s)} = {}^s X_t$, $s = [d_X + 0.5]$, 是一均值为零协方差平稳的随机过程, 其谱密度函数满足下式成立

$$f_U^{(s)}(\cdot) \sim G^{-2(d_X + s)}, \quad 0^+ \quad (2)$$

且 $d_X \in [-0.5, 0.5)$, 其中 $[d_X + 0.5]$ 表示小于或等于 $d_X + 0.5$ 的最大整数. G 为一正的有限常数.

一般的, 若时间序列 $\{Y_t\}$ 具有趋势, 则可表示为一非平稳随机过程 X_t 和 q 阶趋势项 T_t 两部分的和, 即

$$Y_t = X_t + T_t \quad T_t = \sum_{r=0}^q t^r \quad (3)$$

式中: Y_t 和 X_t 具有相同的长记忆参数 ($d_Y = d_X$). 尽管非平稳随机过程的谱密度函数并不存在, 但可定义伪谱密度函数

$$f_X(\cdot) = |1 - \exp(i\cdot)|^{-2s} \\ f_U^{(s)}(\cdot) \sim G^{-2d_X}, \quad 0^+ \quad (4)$$

其中: $f_U^{(s)}(\cdot)$ 满足式 (2), 且当 $d_X = -0.5$ 时, $f_X(\cdot)$ 为谱密度函数, 注意当 $d_X > 0.5$ 时其不是谱密度函数.

1.3 频域分析与数据加窗处理

时间序列在频域中基本的统计方法为离散傅立叶变换和周期图方法. 对时间序列 $X_t, t = 1, 2, \dots, n$, n 为样本容量, 在频率 ω 处的离散傅立叶变换 (以下简称为 DFT) 为

$$w_X(\omega) = (2/n)^{-1/2} \sum_{t=1}^n X_t \exp(i\omega t) \quad (5)$$

其中: $\omega = 2\pi j/n, j$ 为整数. 序列 X_t 在频率 ω_j 处的周期图可表示为 $I_X(\omega_j) = |w_X(\omega_j)|^2$, 对平稳过程, 周期图为谱密度函数的渐进无偏估计量.

然而不幸的是, 周期图虽然是谱密度函数的渐进无偏估计, 但却不是一致估计, 因此如何降低周期图的方差从而将其改造成为一种一致估计就尤为重要. 此外由于序列样本的有限性, 基于周期图的估计会遭到由旁瓣泄漏引起的偏差, 且当谱密度函数在某些频率有峰值时这种偏差将加剧. Tukey^[13] 指出可以通过将原数据乘上一个适当的窗函数而使这两个问题得到解决, 加窗的谱密度函数可表示为

$$\tilde{I}_X(\omega_j) = \left(2 \sum_{t=1}^n h_t^2\right)^{-1/2} \sum_{t=1}^n h_t X_t \exp(i\omega_j t) \quad (6)$$

其中: $\{h_t\}_{t=1}^n$ 为窗函数, 显然, 当 $h_t = 1$ 时, $\tilde{I}_X(\omega_j)$ 即为 $I_X(\omega_j)$. 频率 ω_j 处的加窗周期图为 $\tilde{I}_X(\omega_j) = |\tilde{w}_X(\omega_j)|^2$.

有多种数据窗函数可以使用, p 为正整数, 表示数据窗的平滑程度. $p = 1, h_t = 1$, 相当于没有进行数据加窗处理; $p = 2$, 窗函数为 Bartlett 窗或

triangular 窗; $p = 3$, 为 Kolmogorov-Zhurbenko 窗等. 当 $p > 3$ 时, 窗函数比较复杂, 但随着计算机技术的迅猛发展, 应用傅立叶变换, 窗函数的计算已不再困难.

此外实际应用中, p 需满足以下两个条件: 1) $p > s + 1$, s 如前定义; 2) $p > q + 1$, q 表示序列中存在的确定性趋势的阶数. 实践表明, 平稳过程谱密度函数的加窗周期图估计的方差将随 p 的增大而减小, 且由旁瓣泄漏引起的偏差也会显著减小. 也就是说, 只要应用足够高阶的窗函数, 以上提到的两个问题便会得到控制. 这一性质同样可以扩展到非平稳过程. 虽然非平稳过程没有谱密度函数, 但当 $d_x < 0.5$ 时由式(4)定义的伪谱密度函数是其加窗周期图的极限期望. 事实上, 若序列中不存在确定性趋势, Velasco 证明了只要 $p > d$, 则用 Robinson^[14~17] 介绍的半参数方法进行估计, 所得到的长期记忆参数 d 均为一致估计. 然而若周期图没有进行数据加窗平滑处理, 则估计得到的 d 值仅在 $(-0.5, 0.5)$ 范围内才是一致估计.

若序列中存在趋势, 对于非平稳随机过程 Y_t , 如式(3)所示可写为两部分的和. 估计其长期记忆参数, 通常的, 分两步骤进行: 首先去掉其确定性趋势 T_t , 然后基于去掉趋势的序列估计其长期记忆参数. 去趋势方法一般有两种: 第 1 种对原始序列估计确定性趋势, 然后从中抽取这一趋势; 第 2 种方法即对原始序列进行差分. 这两种方法都有很大的局限性: 对于第 1 种方法, 基于残差序列 $Y_t - T_t$ 估计的长期记忆参数 d_Y 并不一定是一致估计, 因为去掉的拟合趋势项 T_t 可能正好与待估参数 d_Y 有很强的相关性; 对于第 2 种差分方法虽然避免了趋势估计的问题, 但容易产生过度差分或差分不足, 过度差分会导致序列的不可逆性, 而差分不足会导致序列的不平稳性. 因此本文所介绍的加窗周期图估计方法是对原始序列进行直接估计, 避免了以上两种方法的弊端, 且保证了估计结果的一致有效性.

当然, 若 $d_Y < 0.5$, 即满足平稳性要求, 则没有必要进行数据加窗处理, 因为用 Robinson 提出的方法进行的估计本身即为一一致估计. 但 Velasco 证明了在这种情况下采用加窗处理, 结果仍是一致的, 且能够降低由于样本量的有限性而导致的偏差问题.

1.4 多变量随机过程中长期记忆性参数的两步半参数估计方法

下面讨论非平稳向量随机过程的长期记忆性. 设一实值 N 维随机过程 $X_t, t = 0, \pm 1, \dots$, 具有长期记忆性, 记忆参数为 $d^0 = (d_1^0, d_2^0, \dots, d_N^0)$, 其中: 对 $\forall j, 1 \leq j \leq N$, 都有 $d_j^0 > -0.5$, 即满足可逆性. 记 $s^0 = \lfloor d^0 + 0.5 \rfloor$. 定义均值为 μ 的协方差平稳的 N 维向量随机过程 $U_t = \text{diag}\{s^0\} X_t$, 其滞后 j 期的自协方差矩阵为 $\gamma_j = E[(U_t - \mu)(U_{t+j} - \mu)']$. 设其谱密度矩阵 $f(\cdot)$ 存在, 则与 γ_j 有如下关系成立

$$\gamma_j = \int_{-\pi}^{\pi} f(\lambda) e^{ij\lambda} d\lambda$$

与单变量序列的谱密度函数可表示为式(1)的形式一样, 多变量随机过程的谱密度函数 $f(\cdot)$ 可表示为式(7)的形式

$$f(\lambda) \sim R^0 R^+ \lambda^{-d^0}, \quad 0 < \lambda < \pi \quad (7)$$

其中: R^0 为对称、正定实矩阵, 且有 $R^0 = \text{diag}\{s^0 - d^0\}$, $-0.5 < s^0 - d^0 < 0.5$.

对长期记忆参数 d 的估计, 本文介绍的是基于 Whittle 对数似然函数的频域半参数方法, 和其他半参数估计方法相比, 它和拟极大似然估计(QMLE)一样具有许多优点: 如估计结果是渐进正态的, 且不必假定高斯增量性; 估计过程中, 用户不必选择一些可能对结果产生影响的参数; 估计结果更加有效等. 由于省去了多维求极值问题, 因而在计算上又远比 QMLE 方法简单.

目标函数为

$$L_p(R, d) = \frac{d}{m} \sum_{j(p)}^m \log | \gamma_j R^{-1} | + \frac{d}{m} \sum_{j(p)}^m \text{tr} [(\gamma_j R^{-1})^{-1} I^T(\gamma_j)] \quad (8)$$

其中: m/p 为整数, $j(p)$ 表示 $j = p, 2p, \dots, m$, $\gamma_j = \text{diag}\{ \gamma_j^d \}$, $I^T(\gamma_j) = \text{Tr}(\gamma_j) \text{Tr}(\gamma_j)^*$, $I^T(\gamma_j)$ 为加窗周期图, $\text{Tr}(\gamma_j)$ 定义见式(6), $*$ 为共轭符号. 将式(8)两边分别对 R 求偏导, 再令其等于零, 得到如下形式的解

$$R_p(d) = \frac{d}{m} \sum_{j(p)}^m \{ \gamma_j^{-1} \text{Re} \{ I^T(\gamma_j) \} \gamma_j^{-1} \} \quad (9)$$

将式(9)代入式(8), 则目标函数可被简化为

$$p(d) = \frac{-2d}{m} \sum_{j(p)}^m d \log(\gamma_j) + \log | R_p(d) |$$

基于此目标函数的两步半参数估计方法(以下简称 TSE 方法), 估计步骤为: 第 1 步应用 Robinson

的 QMLE 单变量方法估计每个分量序列的长期记忆参数 $\alpha^{(1)}$; 第 2 步应用下面的表达式计算 $\alpha^{(2)}$

$$\alpha^{(2)} = \alpha^{(1)} - \left(\frac{\partial^2_p(d)}{\partial d \partial d} \Big|_{\alpha^{(1)}} \right)^{-1} \cdot \left(\frac{\partial_p(d)}{\partial d} \Big|_{\alpha^{(1)}} \right) \quad (10)$$

当然, 第 1 步不一定非用 QMLE 方法估计 $\alpha^{(1)}$, 亦可用 Robinson 提出的平均周期图等其他方法估计, 最终结果不受影响。

以上的 TSE 估计方法是渐进正态的, $\alpha^{(2)}$ 满足

$$\sqrt{m}(\alpha^{(2)} - d^0) \xrightarrow{d} N(0, p\phi_p E^{-1}) \quad (11)$$

其中: d^0 表示真实的长期记忆参数, $E = 2(I_N + (R^0 \circ R^{0^{-1}}))$, $R = R_p(\alpha^{(2)})$ 为 R^0 的一致估计量, \circ 表示两个矩阵的 Hadamard 乘积,

$$\phi_p = \lim_n \left(\sum_{t=1}^n h_t^2 \right)^{-2} \sum_{k=0, p, 2p, \dots}^{n-p} \left(\sum_{t=1}^n h_t^2 \cos t k \right)^2$$

有关的理论证明请参见 Lobato 和 Velasco。

考虑零假设: 关于 d^0 的 r 个相互独立的约束的线性组合 $Hd^0 = z$, H 为 $r \times N$ 维矩阵, z 为 $r \times 1$ 维向量。则统计量

$$T = \frac{m}{p} \phi_p^{-1} (H\alpha^{(2)} - z) (HE^{-1}H)^{-1} (H\alpha^{(2)} - z) \quad (12)$$

在零假设下服从渐进 χ^2_r 分布。应用式(12) 可检验向量的各分量序列是否具有相同的长期记忆参数。此时, z 为 $N-1$ 维的零向量, $H = (I_{N-1}, 0) - (0, I_{N-1})$, 为 $(N-1) \times N$ 维矩阵, I_{N-1} 为 $N-1$ 维单位矩阵, 0 为 $N \times 1$ 维零向量, N 为样本容量。在文章第 2 部分, 作者将用式(12) 检验中国股票市场波动性序列和交易量序列是否具有相同的长期记忆性。即冲击对这两个序列的影响是否以相同的双曲率衰减。

2 实证研究

2.1 样本数据、来源及数据处理

应用上面介绍的方法, 本文考察了中国股票

表 1 邯钢钢铁与深发展 A 日数据的基本统计量

Table 1 Summary Statistics for Daily Data of Handan Gangtie and Shen Fazhan A

股指	样本	均值	标准差	最小值	最大值	偏度	峰度	$Q(20)$
邯钢钢铁	absR	1.399	1.456	0.000	9.593	2.475	8.534	394.28 ***
	V	1.821	2.873	0.155	31.198	4.580	29.308	2.717 ***
深发展 A	absR	1.661	2.084	0.000	22.64	3.586	20.431	621.25 ***
	V	0.013	0.019	0.001	0.195	3.148	14.034	10.343 ***

*** 表示在 0.01 显著性水平下统计显著

市场的成交量序列和收益波动性序列的长程相关关系, 数据分别为上证 30 指数和深证成分指数所包含股票的每日收盘指数和成交量, 时间跨度为 1995/01/03—2001/03/30。若在 1995 年 1 月 3 日以后上市的公司则以其上市日为样本起始日, 样本容量分别为 718 - 1513(上证) 和 406 - 1510(深证) 个交易日数据, 数据来自“钱龙系统”。股指回报采用对数收益率即 $R_t = \ln(p_t/p_{t-1}) * 100$, 且均进行了分红配股修正, 收益波动性用股指回报的绝对值表示。用换手率作为日成交量序列的替代指标 $V_{j,t} = (S_{j,t}/N_{j,t}) * 100$, 其中 $S_{j,t}$ 表示 j 股票第 t 天的成交量, $N_{j,t}$ 表示 j 股票第 t 天的流通股数目。

中国股票市场自成立以来得到了迅猛地发展, 成交量序列具有明显的增长趋势且并不稳定, 因此依上面的分析, 研究成交量的长期记忆性, 可设数据窗口的阶数为 $p = 2$ 或 $p = 3$, 即假设成交量序列具有一次或二次的确定性趋势。用这种方法, 既省去了估计趋势的麻烦, 又能保证参数估计结果的一致可靠性。

如引言中所说, 有关量价关系的最具权威的经济理论当属混合分布假说(MDH) 模型, 其精髓指出了价格波动性与交易量之间的关系是由到达市场的信息流驱动的。因此, 依据 MDH, 收益波动性序列和交易量序列应具有相同的随机属性。下面文章就将对此进行研究。

2.2 成交量序列与波动性序列各自的长期记忆性

为了叙述上的方便, 下面将分别具体介绍排在上证 30 指数中的第一支股票邯钢钢铁和排在深证成分指数中的第一支股票深发展 A, 样本量分别为 762 和 1 502。在计算过程中, 文章采用两步半参数估计方法, 在 SPLUS 统计语言环境下, 经编程计算, 得到各有关结果。表 1 给出了邯钢钢铁与深发展 A 日回报的绝对值序列与成交量序列的基本统计量。

表1中的 Q 项为 Box-Pierce 统计量, 括号中的数字代表滞后阶数. 由表 1 可知, 邯钢钢铁和深发展 A 的股指回报绝对值序列和成交量序列都具有正的偏度, 且峰值均显著大于 3, 显示了明显的尖峰、厚尾现象; 收益绝对值序列和成交量序列的 Box-Pierce Q 统计量均在 1% 的置信水平上统计显著, 说明两序列均存在明显的序列自相关. 同时也反应了成交量所替代的信息流是序列自相关的.

对于带宽参数 m 的选择, 分别针对 $m_j = (T_j)^{1/2}$ 和 $m_j = 1.5 * (T_j)^{1/2}$ 进行了参数估计 (T_j 为第 j 支股票的样本容量), 估计结果并没有明显区别, 表明 TSE 两步估计方法对 m 具有鲁棒性.

表 2 给出了邯钢钢铁和深发展 A 两支股票的估计结果. 图 1 至图 4 分别给出了上证 30 和深证成分中各股票估计结果的直方图, 文中给出的估计结果均是针对 $p = 1$ 和 $m_j = (T_j)^{1/2}$ 计算所得.

表 2 邯钢钢铁与深发展 A 的估计结果

Table 2 Estimate results of Handan Gangtie and Shen Fazhan A

股指	样本	d	t 统计量	Wald 统计量
邯钢钢铁	absR	0.480 ***	4.120	0.304
	V	0.550 ***	4.716	
深发展 A	absR	0.648 ***	6.736	0.355
	V	0.709 ***	7.382	

注: *** 号表示在 0.01 置信水平下统计显著

由表 2 可知, 邯钢钢铁和深发展 A 的波动性序列和成交量序列的 d 值均显著不同于 0 和 1 (*** 号表示在 0.01 置信水平下统计显著), 表明这两个序列均具有显著的长期记忆性. 注意到成交量序列的 d 值均大于 0.5, 进一步说明了成交量序列的非平稳性.

图 1 描述的上证 30 所有股票收益绝对值的长期记忆参数取值均在 0.019 和 0.634 之间, 均值为 0.318, 中值为 0.304.

图 2 表明个股换手率的长记忆参数取值均在 0.099 和 0.749 之间, 均值为 0.423, 中值为 0.419. 由图 3 可知深圳成分 40 支股票的波动性的长期记忆参数取值均在 0.02 和 0.648 之间, 均值为 0.357, 中值为 0.372; 图 4 表明换手率的 d 值均在 0.089 和 0.733 之间, 均值为 0.484, 中值为 0.507. 且所有这些估计值的 t 统计量均在 5% 的显著性水平下统计显著.

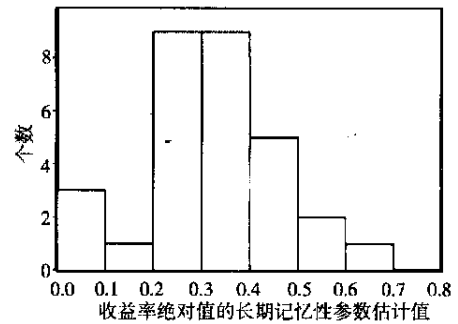


图 1 上证 30 个股长期记忆参数估计结果直方图 / R / Fig. 1 Histogram for 30 estimates of daily absolute returns of Shanghai 30 index

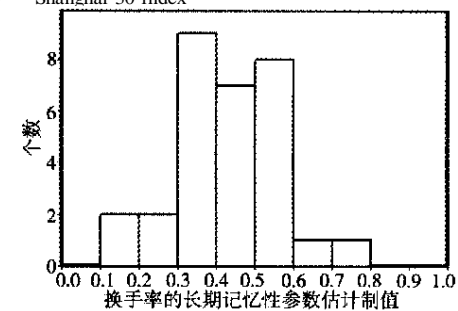


图 2 上证 30 个股长期记忆参数估计结果直方图 V Fig. 2 Histogram for 30 estimates of turnover of Shanghai 30 index

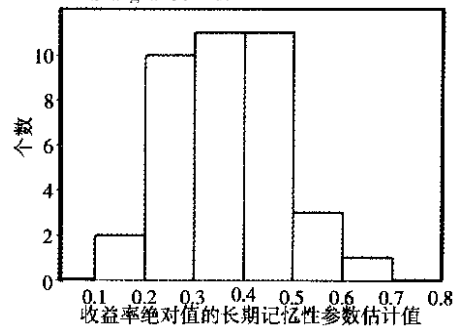


图 3 深圳成分个股长期记忆参数估计结果直方图 / R / Fig. 3 Histogram for 40 estimates of daily absolute returns for Shenzhen component index

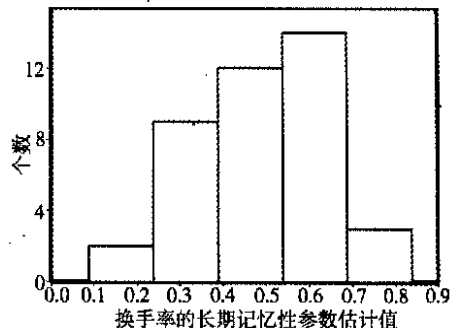


图 4 深圳成分个股长期记忆参数估计结果直方图 V Fig. 4 Histogram for 40 estimates of turnover for Shenzhen component index

2.3 成交量序列与波动性序列的共同的长期记忆性

由上面的分析可知中国股票市场的波动性序列及换手率序列都具有显著的长期记忆性,但这两个序列的长期记忆性是否相同?即它们的长期记忆参数是否相等呢?如果换手率和波动性的长期相关关系的确是由一共同的、潜在的信息到达过程驱动,则这两个序列的长期记忆参数值应相当接近.图 5、图 6 分别给出了上证 30 所包含的 30 支股票和深圳成分所包含的 40 支股票的长期记忆参数散点图 ($d_{j,1}, d_{j,2}$), $d_{j,1}$ 表示第 j 支股票的收益绝对值的长期记忆参数, $d_{j,2}$ 表示第 j 支股票的换手率序列的长期记忆参数.由图可知,尽管两图均具有奇异点,但绝大多数点则与 45 度直线很接近,即符合前面的推测,绝大多数股票的换手率序列和波动性序列均具有相同的长期记忆性参数.

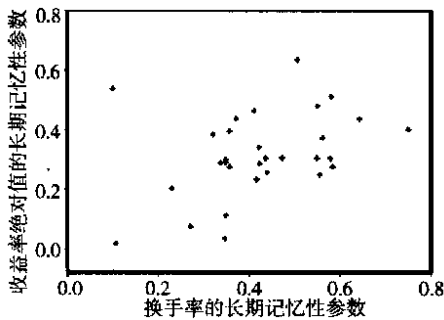


图 5 上海长期记忆性参数估计结果散点图

Fig. 5 Scatterplot for estimate results of the long memory parameter for daily absolute returns and turnover in Shanghai

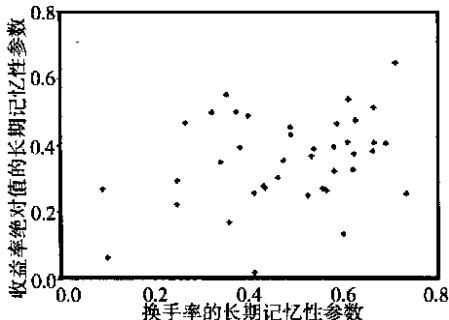


图 6 深圳长期记忆性参数估计结果散点图

Fig. 6 Scatterplot for estimate results of the long memory parameter for daily absolute returns and turnover in Shenzhen

要得到更具有说服力的结论,还要进行正式的统计检验,下面基于式(12)进行 Wald 检验.由于进行的是双变量检验,式(12)可简化为式(13):

$$T = \frac{m}{p} \phi_p^{-1} (H\hat{H}^{(2)}) (HE^{-1}H)^{-1} (H\hat{H}^{(2)}) \quad (13)$$

其中, $H = (1, -1)$, 如上所述, T 统计量在零假设,即两序列具有相同的长期记忆参数下服从渐进 χ^2_1 分布,即使所检验的序列中存在的真实的随机趋势的阶数小于 p 值,结论仍然成立.

如表 2 所示,应用式(13)计算的邯鄯钢铁的 Wald 统计量的值为 0.304, 5% 显著性水平下, (1) 的临界值为 3.841, 因此检验接受零假设 $d_{j,1} = d_{j,2}$; 计算得到的深发展 A 的 Wald 统计量为 0.355, 显然小于 3.841, 因此也接受零假设 $d_{j,1} = d_{j,2}$. 即邯鄯钢铁和深发展 A 两支股票的波动性序列和换手率序列均具有相同的长期记忆性.

图 7、图 8 描述了上证 30 和深圳成分个股的波动性序列与换手率序列具有相同的长期记忆参数的 Wald 检验结果直方图. 上证 30 支股票的 Wald 统计量的值有 76.7% 小于 3.841, 即有 23 支股票接受零假设, 具有相同的长期记忆参数; 深圳成分 40 支股票的 Wald 统计量有 77.5% 的值小于 3.841, 即有 31 支股票接受零假设. 综上所述, 无论是上海股市还是深圳股市均可认为波动性序列和换手率序列具有相同的长期记忆参数.

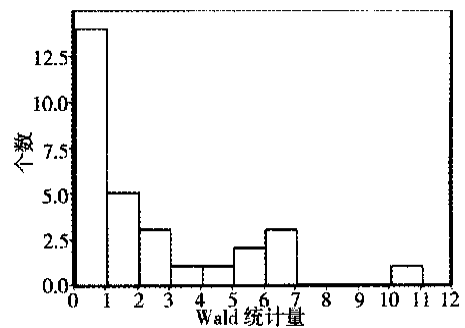


图 7 上海 Wald 统计量直方图

Fig. 7 Histogram for wald statistics of Shanghai 30 index

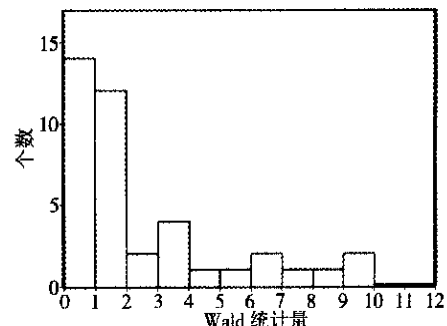


图 8 深圳 Wald 统计量直方图

Fig. 8 Histogram for wald statistics of Shenzhen component index

3 结 论

本文介绍了一种估计多变量非平稳时间序列的长期记忆性参数的新方法:两步半参数估计方法.该方法不必对原始数据估计趋势,也不必先对其进行差分使其成为一平稳时间序列然后再进行参数估计,而是应用加窗处理,直接利用原始数据便可以得到长期记忆性参数的一致有效估计.在此方法的基础上,文章研究探讨了中国股票市场上证 30 指数和深圳成分指数所包含的个股股票的换手率序列和收益绝对值序列的长期相关关系,结果表明中国股票市场的这两个序列均具有

明显的长期记忆性,并且绝大多数股票的波动性序列和成交量序列具有相同的长期记忆参数.因此尽管一冲击对波动性和成交量两序列的影响就短期而言可能是截然不同的,但长期而言,这一冲击的影响却是以相同的速率缓慢衰减.即驱动量价关系的潜在信息流本身亦为一长期记忆过程.

收益波动性序列、成交量序列表现出相同的长期记忆性,为进一步研究二者之间的分数协整性,进而探求二者之间的长期均衡关系提供了可能.然而这一问题虽然早已引起了国际学者的广泛关注,但由于相关渐进分布理论发展的相对滞后,至今仍未被解决,还有待于进一步的研究探讨.当然,这也将是作者下一步的研究目标.

参 考 文 献:

- [1] Engle R F, Bollerslev T. Modeling the persistence of conditional variances[J]. *Econometric Reviews*, 1986, 5: 1—50.
- [2] Baillie R T, Bollerslev T, Mikkelsen H O. Fractionally integrated generalized autoregressive conditional heteroskedasticity[J]. *Journal of Econometrics*, 1996, 74: 3—30.
- [3] Bollerslev T, Mikkelsen H O. Modeling and pricing long memory in stock market volatility[J]. *Journal of Econometrics*, 1996, 73: 151—184.
- [4] 李汉东, 张世英. 自回归条件异方差的持续性研究[J]. *预测*, 2000, 1: 51—54.
Li Han-dong, Zhang Shi-ying. The persistence of autoregressive conditional heteroscedasticity[J]. *Forecasting Journal(Bimonthly)*, 2000, 1: 51—54. (in Chinese)
- [5] 王春峰, 张庆翠. 中国股市波动性过程中的长期记忆性实证研究[J]. *系统工程*, 2004, 1: 78—83.
Wang Chun-feng, Zhang Qing-cui. Modeling long memory in Chinese stock market volatility[J]. *Systems Engineering*, 2004, 1: 78—83. (in Chinese)
- [6] Epps T W, Epps M L. The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis[J]. *Econometrica*, 1976, 44: 305—321.
- [7] Lamoureux C, Lastrapes W D. Endogenous trading volume and momentum in stock return volatility[J]. *Journal of Business & Economic Statistics*, 1994, (12): 253—260.
- [8] Bollerslev T, Jubinski D. Equity trading volume and volatility: Latent information arrivals and common longrun dependencies[J]. *Journal of Business & Economic Statistics*, 1999, 17: 9—21.
- [9] Lobato I N, Velasco C. Long memory in stock market trading volume[J]. *Journal of Business & Economic Statistics*, 2000, 18: 410—427.
- [10] Lobato I N. A semiparametric two step estimator in a multivariate long memory model[J]. *Journal of Econometrics*, 1999, 90: 129—153.
- [11] Velasco C. Non-stationary log-periodogram regression[J]. *Journal of Econometrics*, 1999a, 91: 325—371.
- [12] Velasco C. Gaussian semiparametric estimation of non-stationary time series[J]. *Journal of Time Series Analysis*, 1999b, 20: 87—127.
- [13] Cooley J W, Tukey J W. An algorithm for the machine calculation of complex fourier series[J]. *Mathematics of Computation*, 1965, 19: 267—301.
- [14] Robinson P M. Semiparametric analysis of long memory time series[J]. *The Annals of Statistics*, 1994a, 22: 515—539.
- [15] Robinson P M. Rates of convergence and optimal spectral bandwidth for long range dependence[J]. *Probability Theory and Related*

- Fields, 1994b, 99: 443—473.
- [16] Robinson P M. Log-periodogram regression of time series with long range dependence[J]. The Annals of Statistics, 1995a, 23: 1048—1072.
- [17] Robinson P M. Gaussian semiparametric estimation of long range dependence[J]. The Annals of Statistics, 1995b, 23: 1630—1661.
- [18] Geweke J, Porter H S. The estimation and application of long memory time series models[J]. Journal of Time Series Analysis, 1983, 4: 221—238.
- [19] Andersen T G, Bollerslev T. Heterogeneous information arrivals and return volatility dynamics: Uncovering the longrun in high frequency returns[J]. Journal of Finance, 1997, 52: 975—1005.
- [20] Cooley J W, Tukey J W. An algorithm for the machine calculation of complex fourier series[J]. Mathematics of Computation, 1965, 19: 267—301.

Research on common long memory between trading volume and volatility in Chinese stock market

ZHANG Qing-cui, WANG Chun-feng

Institute of Systems Engineering, Tianjin University, Tianjin 300072, China

Abstract: This article examines the behavior of equity volume and volatility for the individual firms composing the Shanghai 30 index and Shenzhen composite index in long run. Using multivariate frequency domain two-step semi-parametric procedures, by tapering the data instead of detrending them, the long memory parameters of nonstationary vector process have been consistently estimated. We find that there is strong long memory in both series, besides, for most of the stocks, the volatility and volume exhibit the same degree of long memory. Which indicate the latent information-arrival structure which determined the volume-volatility relationship also possess long memory characteristics.

Key words: long memory; nonstationary stochastic process; spectral analysis; tapering periodogram