

文章编号: 1007-6735(2011)05-0433-11

复杂金融系统的重现时间间隔分析

任 飞, 顾高峰, 蒋志强, 周炜星

(华东理工大学 商学院, 上海 200237)

摘要: 依据复杂性科学的思路,研究了金融市场大波动极端事件的重现时间间隔,考察了中国股市高频数据的重现时间间隔分布和时间关联特性,介绍了几种分布检验和关联测度方法,从不同角度进行精确分析.其次,系统介绍了重现时间间隔分析方法在金融复杂系统研究中的应用,对波动率、已实现波动率、收益率和交易量的重现时间间隔进行了实证分析,并在此基础上进行初步市场风险估计.最后,介绍了基于委托驱动的微观模型,通过模拟交易人的委托下单过程模拟市场的价格波动演化,研究大波动极端事件重现时间间隔的动力学机理,为中国股市大波动极端事件的风险估计和规避提供理论依据.

关键词: 金融物理学; 极端事件; 重现时间间隔分析; 金融风险

中图分类号: F 830.91 **文献标志码:** A

Recurrence interval analysis of complex financial systems

REN Fei, GU Gao-feng, JIANG Zhi-qiang, ZHOU Wei-xing

(School of Business, East China University of Science and Technology, Shanghai 200237, China)

Abstract: This paper uses the methods of complex science, and studies the statistical properties of probability distribution and memory effect of the recurrence intervals, defined as the intervals between extreme events in financial markets. Relatively accurate analysis is performed by using various goodness-of-fit tests and memory effect detecting methods. The application of the recurrence interval analysis in financial complex systems is systematically introduced, including the empirical studies of recurrence intervals of volatilities, realized volatilities, returns and trading volumes. Based upon this study, we perform a primarily risk estimation for the stock market. Furthermore, an order driven model is introduced. The dynamics of the recurrence intervals between extreme events is studied by mimicking the order submission process of the investors in stock markets. This work may provide a theoretical foundation for the risk estimation and avoiding of extreme events in the Chinese stock market.

Key words: *econophysics; extreme events; recurrence interval analysis; financial risk*

收稿日期: 2011-09-16

作者简介: 任 飞(1980-),女,副教授.研究方向:金融复杂系统、金融工程、金融风险控制与监管.

E-mail: fren@ecust.edu.cn

近年来我国经历了暴雪、地震、干旱、金融危机等一系列事件,这些异常的、平时不易发生的事件在统计学意义上称为极端事件.肇始于美国次贷危机的全球金融危机,被认为是1929年美国股灾后近百年来最严重的金融危机,当前面临的欧洲债务危机,也可能导致世界经济的再次探底.研究金融极端事件的统计规律和内在动力学机制,对估计金融市场风险并采取措施规避风险有着十分重要的意义.传统经济学理论框架建立在有效市场假设的基础之上,无法解释和预言此次金融危机.当前面临的经济困境促使科学家重新审视主流经济学理论,著名物理学家 Bouchaud 和经济学家 Lux 和 Westerhoff 分别在《Nature》和《Nature Physics》上发表评论,指出现有的经济理论需要根本性的科学革命,应该从实验金融数据出发归纳总结金融市场自身的运动规律^[1],以解决目前的金融危机问题^[2].

金融复杂系统具有和自然界复杂系统十分类似的特征行为,受此启发人们运用复杂性科学、非线性科学和计算实验科学等自然科学的方法来对其进行研究^[3-6].金融市场内伴随某个大波动极端事件,紧接着发生一系列后续波动,同多粒子体系的自组织临界现象非常相似.诱发后续波动的大波动极端事件决定了金融市场的动力学行为及其复杂性,直接反映了市场的风险水平,采用重现时间间隔的分析方法研究大波动,能为金融市场的风险管理提供新的视角和思路.

1 基本概念、方法、理论结果

重现时间间隔(recurrence interval)定义为某个连续事件之间的时间间隔,早期集中在对自然界极端事件的研究中,比如洪水、高温天气和地震等自然灾害^[7-9].金融极端大波动事件的重现时间间隔研究已经引起了许多学者的关注^[10-16].股票价格的波动率或收益率连续地超过某个阈值 q 的大波动事件间的时间间隔,简称为重现时间间隔 τ .可根据重现时间间隔的概率分布特征,由较大阈值的时间间隔分布推理得到极端大波动事件的时间间隔分布规律,能克服极端大波动统计样本不足的困难,对于金融市场的风险估计和管理具有现实意义.进一步的实证和模拟研究表明,重现时间间隔的这种标度行为可能产生于原始记录的长期记忆效应(long-term memory)^[12-15].

关于大波动重现时间间隔的分布形式还没有统

一定论.早期大波动的重现时间间隔分布研究主要是基于概率分布的定性观测.Yamasaki 等^[10]首次运用重现时间间隔的分析方法研究股票的波动率,采用美国股市的日度数据,将波动率简单定义为对数收益的绝对值,发现波动率重现时间间隔的概率密度具有标度行为(scaling behavior).Wang 等^[11]采用美国股市的日内高频数据,证实了波动率重现时间间隔分布的标度行为,其具体分布形式遵从拉伸指数分布(stretched exponential distribution),Jung 等^[16]在日本股市中也观测到类似的拉伸指数分布.

随着研究深入到对高频数据的分析,一系列高精度的定量分析方法被引入,例如 KS(kolmogorov-smirnov)检验方法、 m -阶矩测量、基于 bootstrap 方法的 KS 拟合度检验等.Wang 等^[17]采用纽约证交所的逐笔交易与报价数据,通过测量 $\tau/\langle\tau\rangle$ 的 m -阶矩,发现间隔时间的分布略微偏离标度律,类似的偏离标度律行为在韩国股市中也被观测到^[18].Bogachev 等^[19]也发现重现时间间隔的分布随阈值 q 变化,且分布形式遵循幂律分布.

具有多体相互作用的复杂体系具有长程时间关联(也称为记忆效应),是构造极端事件的重现时间间隔分布模型的主要依据.经验数据研究表明,极端事件的间隔时间分布与波动率的长程关联密切相关^[12-15],根据这一机理,构造了一系列微观动力学模型.用傅立叶变换构造具有分形和线性长程关联特性的时间序列,其间隔时间分布具有标度行为,且分布服从拉伸指数分布^[4,20],其解析解近似可解^[12-13].用多重分形随机行走模型,构造具有多重分形和长程关联特性的时间序列,无论关联为线性或非线性,间隔时间分布都无标度行为,但分布遵循幂律分布^[19,21].

目前,中国股市波动的重现时间间隔研究还比较少.Qiu 等^[22]采用比较直观的方法观测上交所 4 只股票波动率的重现时间间隔的概率分布,认为重现时间间隔的分布具有标度行为,且标度函数服从拉伸指数分布.Ren 等^[23]分析上交所和深交所 30 只交易活跃的股票两年的日内高频数据,采用 m -阶矩测量方法考察波动率重现时间间隔分布的标度行为,发现大多数股票的时间间隔分布偏离标度律,并采用基于 bootstrap 方法的 KS 拟合优度检验证实标度函数遵循拉伸指数分布.最近,Ren 等^[24]采用 20 只在上交所和深交所上市股票的将近 10 年的日内数据,发现收益率的重现时间间隔分布遵循幂律分布,类似结果在交易量的重现时间间隔分析

中也被观测到^[25-26].

1.1 定义

假设有某一股票的时间序列,根据该时间序列可以求得原始变量 $R(t)$, $R(t)$ 可以是波动率、收益率等任一变量.在数据预处理阶段需要对 $R(t)$ 进行归一化处理(normalization)

$$r(t) = \frac{R(t)}{\sqrt{\langle R(t)^2 \rangle - \langle R(t) \rangle^2}} \quad (1)$$

归一化变量 $r(t)$ 的重现时间间隔定义为该变量连续地超过某个阈 q 的时间间隔,用字母 τ 表示.图1所示的是归一化变量 $r(t)$ 的重现时间间隔的示意图,对应波动率和交易量的归一化变量, τ 为由 $q > 0$ 的阈值截取的时间间隔,对于收益率的归一化变量,由于收益率可正可负, τ 为由 $q > 0$ 和 $q < 0$ 截取的时间间隔.每取一个阈值 q ,便可得到阈值下的一组间隔时间序列 τ .可通过观测不同阈值下的重现时间间隔,研究这些超过阈值 q 的事件重现的规律.

1.2 分布及检验方法

经验研究发现重现时间间隔的概率分布可能满足拉伸指数分布、幂律分布或者两者的混合分布,因此需要采用相对精确的拟合和拟合优度检验方法来确定具体的分布形式.常用的拟合方法是由 Newman^[27] 提出的基于 KS 统计量的极大似然估计方法,该方法最早是针对幂律分布的拟合提出的,但该方法对于其它分布的拟合也有较好的适用性.

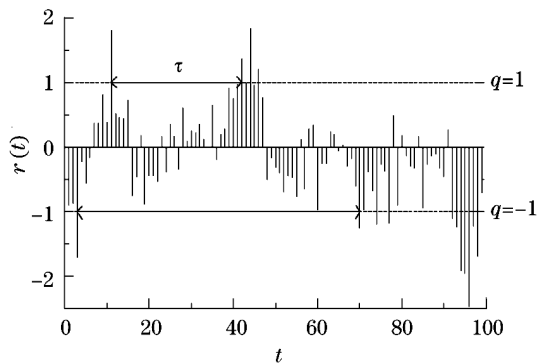


图1 归一化变量 $r(t)$ 的重现时间间隔的示意图

Fig.1 Illustration of recurrence intervals of the normalized returns $r(t)$

某个变量 x 的概率密度函数为 $P(x)$, 其累积概率分布函数为 $F(x)$, 假设变量 x 大于 $\hat{x}_{\min}$ 部分满足某个理论分布, 该理论分布的累积概率分布函数为 $F_{\text{fit}}(x)$, 则 KS 统计量定义为

$$KS = \max_{x \geq \hat{x}_{\min}} (|F - F_{\text{fit}}|) \quad (2)$$

为了使得拟合的理论分布更接近于变量 x 的实际分布, KS 统计量应该尽量的小. $\hat{x}_{\min}$ 值在 KS 统计量取最小值求得, 拟合分布的参数由极大似然估计解得.

对于已知分布的检验, 可采用基于 bootstrap 方法的 KS 拟合优度检验, 用来检验实证数据是否服从拟合得到的理论分布. 前面已经求得变量 x 的实证数据的 KS 统计量, 我们还可以计算加权的 KS 统计量

$$KSW = \max_{x \geq \hat{x}_{\min}} \left(\frac{|F - F_{\text{fit}}|}{\sqrt{F_{\text{fit}}(1 - F_{\text{fit}})}} \right) \quad (3)$$

加权的 KS 统计量 KSW 在累积概率密度函数的边缘会更敏感.

接下来采用 bootstrap 方法随机产生 1 000 组模拟数据, 模拟数据的概率分布满足拟合得到的理论分布, 且模拟数据的数量等于变量 x 的实证数据的样本数. 对每一组生成的模拟数据重构其累积分布函数 F_{sim} , 并且采用 Newman 等人的方法对模拟数据进行拟合, 计算相应的累积分布函数 $F_{\text{sim, fit}}$. 对这 1 000 组模拟数据计算其 KS 统计量和 KSW 统计量

$$KS_{\text{sim}} = \max(|F_{\text{sim}} - F_{\text{sim, fit}}|) \quad (4)$$

$$KSW_{\text{sim}} = \max \left(\frac{|F_{\text{sim}} - F_{\text{sim, fit}}|}{\sqrt{F_{\text{sim, fit}}(1 - F_{\text{sim, fit}})}} \right) \quad (5)$$

KS 拟合优度检验的 p -value 由实证数据的 KS (或 KSW) 统计量和模拟数据的 KS_{sim} (或 KSW_{sim}) 统计量比较生成, 定义为 $KS_{\text{sim}} > KS$ (或 $KSW_{\text{sim}} > KSW$) 的概率. p -value 可以解释为实证数据的分布与拟合得到的理论分布的累积概率分布函数差值小于随机误差 (随机数据的累积概率分布函数差值) 的概率, 也可以理解为实证数据分布服从于拟合分布的概率. 这样就可以通过 p -value 来描述拟合得到的理论分布的拟合优度. 在显著水平为 5% 的条件下, 若 p -value 大于 5%, 则原假设被接受, 我们就可以说重现时间间隔的分布能很好地被理论分布拟合; 若 p -value 小于 5%, 那么拒绝原假设, 说明重现时间间隔的分布不能被理论分布拟合.

1.3 短期记忆性

概率分布并不能完整地描述重现时间间隔序列的所有统计性质, 一个序列的记忆性是另一个独立于其概率分布的重要的统计性质. 若某个变量的序列具有记忆效应, 则该序列中后面的变量值和前面的变量值有一定的相关性. 实证研究已经发现金融

市场的重现时间间隔具有时间记忆效应^[10-13],即处在后面的重现时间间隔和前面的实际间隔有一定关系.

根据记忆时间的长短,记忆性分为短期记忆性和长期记忆性,分别对应于物理学中的短程关联和长程关联.短期记忆性是指,序列中某个时刻的变量值和前面离它不远的变量值有一定的相关性.观测重现时间间隔的短期记忆性,可通过计算重现时间间隔的条件概率分布 $P(\tau | \tau_0)$,即紧接着时间间隔 τ_0 之后出现时间间隔 τ 的概率.在大波动的重现时间间隔研究中,该变量和市场风险估计密切相关,表示前一个大波动事件距离当前大波动事件的时间间隔为 τ_0 ,下一个大波动事件在时间 τ 之后发生的概率.

1.4 长期记忆性

重现时间间隔不仅具有短时记忆性,还具有长期记忆性,即某个时刻的重现时间间隔和较长时间前的重现时间间隔具有相关性.DFA(detrend fluctuation analysis)方法是研究时间序列长期记忆性的一种常用方法,DFA方法被广泛应用于分形和多重分形的研究^[28-29],其二阶的降趋波动函数(detrended fluctuation function)可用来测量时间序列的记忆性^[30].相比较于其它方法,DFA方法对于不稳定的时间序列也适用,可以避免因时间序列不稳定而引起的伪相关性的干扰.

假设存在一个时间序列 $N_i (i = 1, 2, \dots, N)$,将该时间序列进行如下累加处理

$$y(j) = \sum_{i=1}^j [N_i - \langle N_i \rangle] \quad (6)$$

其中 $\langle N_i \rangle$ 为序列的均值.接下来将时间序列 N_i 分割成 K 个不重叠的盒子 $I_n, n = 1, 2, \dots, K$,每个盒子都有相等的窗口长度 s .每个盒子中,用一个多项式 y_{pol} 对累加后的时间序列进行拟合,然后对该序列进行降趋(detrend)

$$Y(j) = y(j) - y_{pol}(j) \quad (7)$$

对于一个给定的时间窗口长度 s ,计算其二阶降趋波动函数(detrended fluctuation function)

$$F(s) = \sqrt{\frac{1}{N} \sum_{j=1}^N [Y(j)]^2} \quad (8)$$

对不同的时间窗口长度 s 都重复上述运算,就可以得到不同 s 值下的降趋波动函数 $F(s)$. $F(s)$ 和 s 存在标度关系(scaling relation)

$$F(s) \propto s^\alpha \quad (9)$$

参数 α 称为 Hurst 指数,该指数取值表征序列的记

忆特性:若 $\alpha = 0.5$,则序列没有相关性;若 $\alpha > 0.5$,则序列存在正相关性;若 $\alpha < 0.5$,则序列存在负相关性^[33].

2 实证分析

2.1 波动率(volatility)

首先采用 30 只在深交所和上交所上市交易的股票的 1 分钟数据,数据的时间跨度为 2004 年 1 月到 2006 年 6 月,分析中国股市价格波动率的重现时间间隔^[23,31].波动率 $R(t)$ 定义为价格对数收益的绝对值

$$R(t) = |\ln Y(t) - \ln Y(t-1)| \quad (10)$$

式中 $Y(t)$ 为第 t 分钟股票的价格.

波动率具有日内效应(intraday pattern),即接近开盘和收盘的时候,波动率较大.可定义日内平均波动率

$$A(s) = \frac{\sum_{i=1}^N R^i(s)}{N} \quad (11)$$

式中 $R^i(s)$ 表示第 i 个交易日日内 s 时刻的波动率.

如图 2 所示,日内平均波动率 $A(s)$ 呈现“U”型形状.

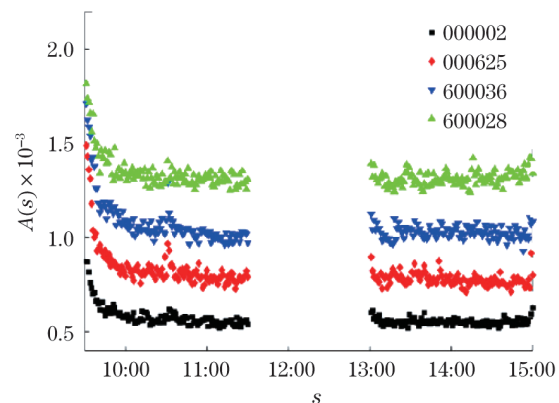


图2 上海和深圳证券交易所内个股的日内平均波动率

Fig.2 Mean intraday volatility of four representative stocks in Shanghai and Shenzhen Stock Exchanges

波动率的日内效应对波动率的重现时间间隔测量有一定影响,会引起以天为周期的周期性效应.因此,在测量波动率的重现时间间隔之前,先要消除日内效应

$$R'(t) = \frac{R(t)}{A(s)} \quad (12)$$

然后,再对消除日内效益的波动率做归一化处理

$$r(t) = \frac{R'(t)}{\sqrt{\langle R'(t)^2 \rangle - \langle R'(t) \rangle^2}} \quad (13)$$

采用 1.1 节的定义,计算归一化波动率在不同阈值 q 下的重现时间间隔 τ ,观测重现时间间隔 τ 的概率分布. 如图 3 所示的是两只股票(000839 和 000625)的重现时间间隔的累积概率分布,可以看出 000839 这只股票不同阈值下的分布不能很好地重合在一条线上,而 000625 这只股票不同阈值下的分布能较好地重合在一条线上,具有标度行为

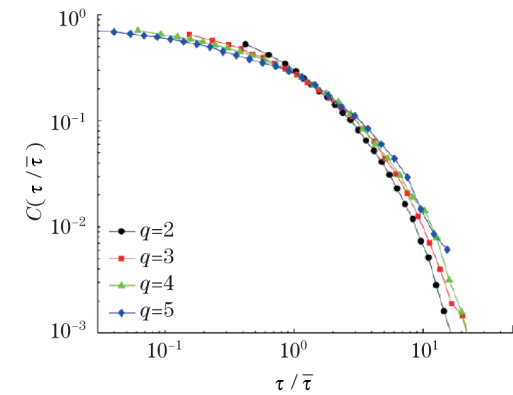
$$P(\tau) = \frac{1}{\langle \tau \rangle} f(\tau / \langle \tau \rangle) \quad (14)$$

式中 $\langle \tau \rangle$ 为平均重现时间间隔.

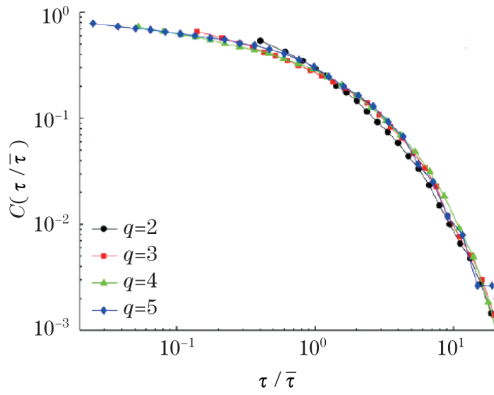
在 30 只个股中,找到 12 只个股具有较好的标度行为,假设其分布满足拉伸指数分布

$$f(\tau / \langle \tau \rangle) = f(x) = ce^{-ax^\gamma} \quad (15)$$

将 $q=2$ 和 $q=5$ 时标度后的重现时间间隔 $\tau / \langle \tau \rangle$ 合并,用最小二乘法拟合两者共同的拉伸指数分布,然后采用 1.2 节介绍的拟合优度检验方法验证以上假设,即 $q=2$ 和 $q=5$ 时, $\tau / \langle \tau \rangle$ 的分布都符合它们共同的拉伸指数拟合分布. 根据表 1 的结果,同时采用 KS 和 KSW 统计量有 6 只股票在 5% 的显著性水平下通过检验,即它们的重现时间间隔分布具有标度行为,且分布服从拉伸指数分布. 关于重现时间间隔分布标度行为的验证,还可以采用重现时间间隔的 m -阶矩测量方法^[21].



(a) 股票000839



(b) 股票000625

图 3 个股 000839 和 000625 波动率重现时间间隔的累积概率分布

Fig.3 Complementary cumulative distribution of the volatility return intervals for stock 000839 and stock 000625

表 1 上海和深圳证券交易所内个股的 KS 拟合优度检验结果

Tab.1 KS-goodness-of-fit test for the individual stocks in Shanghai and Shenzhen Stock Exchanges

股票代码	q	p_{ks}	p_{ksw}	股票代码	q	p_{ks}	p_{ksw}
000625	2	0.224	0.769	600019	2	0.938	0.742
	5	0.045	0.023		5	0.567	0.574
600026	2	0.010	0.215	600036	2	0.311	0.432
	5	0.021	0.011		5	0.086	0.098
600100	2	0.003	0.022	600104	2	0.851	0.973
	5	0.157	0.184		5	0.709	0.866
600171	2	0.158	0.286	600428	2	0.004	0.001
	5	0.401	0.346		5	0.002	0.007
600550	2	0.226	0.354	600601	2	0	0
	5	0.059	0.103		5	0.883	0.130
600602	2	0.995	0.943	600688	2	0.038	0.386
	5	0.996	0.592		5	0.188	0.056

另外,我们还发现波动率的重现时间间隔同时具有短期记忆性和长期记忆性. 通过计算条件概率分布,可观测重现时间间隔的短期记忆性. 假设一只

股票的回复事件间隔序列有 N 个记录,先将各个记录按升序排列后,平均分割成 4 个子集,记为 Q_1, Q_2, Q_3, Q_4 ,最小的 $1/4$ 回复时间间隔记录在 Q_1

中,而最大的 $1/4$ 回复时间间隔记录在 Q_4 中,分别计算 τ_0 落在 Q_1 和 Q_4 这两个子集里的条件概率密度函数 $P_q(\tau|\tau_0)$.

图4给出了股票 000002 以 $\tau/\langle\tau\rangle$ 为自变量,条件概率密度函数 $P_q(\tau|\tau_0)$ 为因变量的图像,空心的符号表示 τ_0 来自子集 Q_4 ,即重现时间间隔最大的子集,实心的符号表示 τ_0 来自子集 Q_1 ,即重现时间间隔最小的子集.可以观测到,对于较大的 $\tau/\langle\tau\rangle$, $\tau_0 \in Q_4$ 的 $P_q(\tau|\tau_0)$ 要比 $\tau_0 \in Q_1$ 的 $P_q(\tau|\tau_0)$ 大,而对于较小的 $\tau/\langle\tau\rangle$, $\tau_0 \in Q_4$ 的 $P_q(\tau|\tau_0)$ 要比 $\tau_0 \in Q_1$ 的 $P_q(\tau|\tau_0)$ 小,在其它股票中也观测到类似的现象.说明较小的重现时间间隔 τ_0 之后更趋向于出现一个小的重现时间间隔,而大的重现时间间隔 τ_0 之后更趋向于出现一个大的重现时间间隔,该结果表明波动率的回复时间间隔具有短期记忆性.

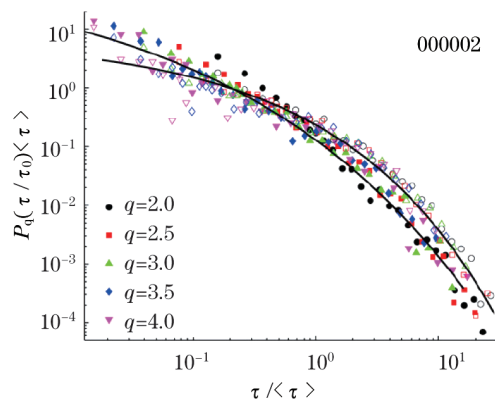


图4 股票 000002 的波动率重现时间间隔的条件概率密度函数

Fig.4 Conditional probability distribution function of the scaled return intervals for a representative stock 000002

我们还采用 1.4 节介绍的 DFA 方法,来研究波动率的重现时间间隔的长期记忆性.图5给出了 30 只股票的 Hurst 指数,圆圈表示的是波动率的重现时间间隔的 Hurst 指数,所有股票的 Hurst 指数都远远大于 0.5,说明波动率的重现时间间隔序列具有长期记忆性.方块表示的是将波动率序列随机打乱后,产生的重现时间间隔序列的 Hurst 指数,大都落到了 $H=0.5$ 这条线附近,长期记忆性被去除了.该结果表明重现时间间隔序列的长期记忆性可能来源于原始波动率序列的记忆效应.

2.2 已实现波动率(realized volatility)

上文介绍了波动率的重现时间间隔,采用对数收益的绝对值定义波动率,是金融物理学中比较常

见的定义方法.但波动率的定义多种多样,研究不同定义下的波动率重现时间间隔的普适性和非普适性特征具有一定意义.在主流经济学研究中,通常用一段时间内的价格方差来度量波动率,例如 Anderson 等^[32]采用日内价格对数收益的平方和定义日度的已实现波动率.借鉴该定义,将 t 到 $t-1$ 时间内所有交易的对数收益率的绝对值之和定义为已实现波动率

$$R(t) = \sum_{t-1 < t' < t} |\ln Y(t') - \ln Y(t'-1)| \quad (16)$$

式中 $Y(t')$ 为 t 到 $t-1$ 时间内 t' 时刻发生的交易的成交价.

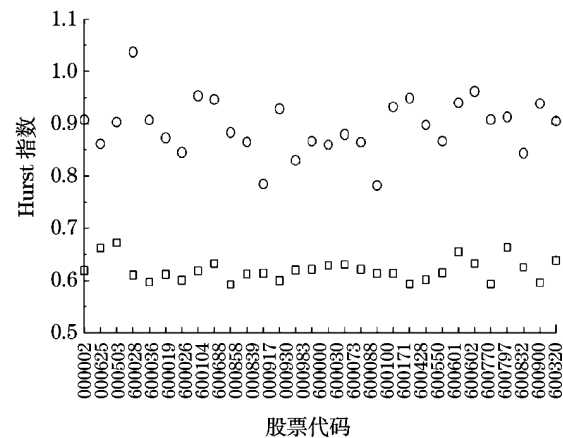


图5 上海和深圳证券交易所内个股的波动率重现时间间隔的 Hurst 指数

Fig.5 Hurst exponents for large scales (circles) and small scales (squares) for the stocks in Shanghai and Shenzhen Stock Exchanges

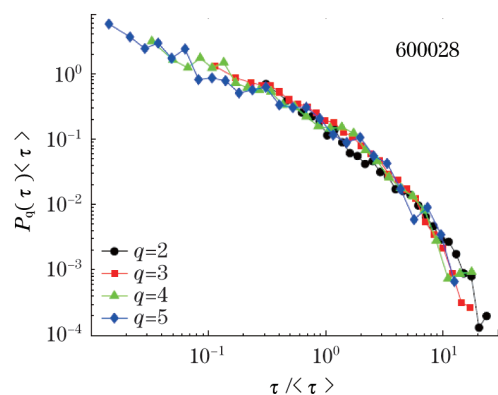
采用在上海证券交易所上市的 22 只股票 2004 年 1 月到 2006 年 6 月的日内超高频数据,该数据包含了这 22 只股票所有分笔交易的信息,在此基础上构造了 1 分钟的已实现波动率,即每分钟内所有交易的对数收益率的绝对值之和.图6比较了采用前一章定义的波动率和已实现波动率的重现时间间隔分布,发现不同阈值 q 下已实现波动率的重现时间间隔分布重合性更好,说明已实现波动率的重现时间间隔分布具有更好的标度行为.进一步研究表明,部分股票的已实现波动率的重现时间间隔分布满足拉伸指数分布,能通过 KS 拟合优度检验,并且已实现波动率的重现时间间隔也具有短期和长期记忆性^[33].

2.3 收益率(return)

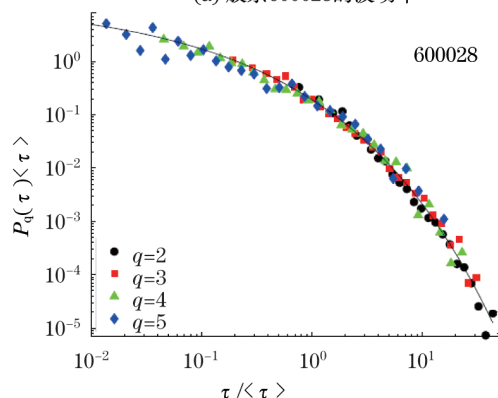
股票的收益率直接关系到价格的涨跌情况,因此研究收益率的重现时间间隔,对股票市场的风险估计和管理研究具有比较重要的意义.采用在上交

所和深交所上市的 20 只个股 2000 年 1 月至 2009 年 5 月、上证综指和深证综指 2003 年 1 月至 2009 年 4 月的 1 分钟数据,研究中国股市收益率的重现时间间隔^[24]. 收益率 $R(t)$ 定义为价格的对数收益, $Y(t)$ 为第 t 分钟股票或指数的价格,则

$$R(t) = \ln Y(t) - \ln Y(t-1) \quad (17)$$



(a) 股票600028的波动率



(b) 股票600028的已实现波动率

图6 股票 600028 的波动率(左)和已实现波动率(右)的重现时间间隔的概率分布

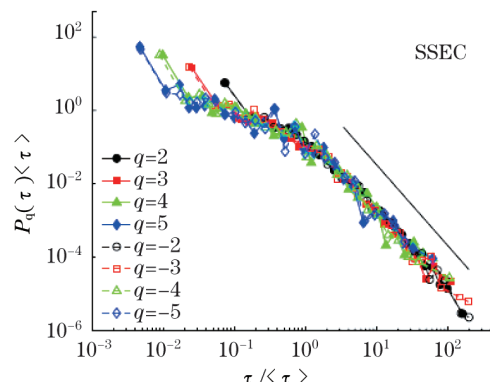
Fig.6 Probability distributions of the return intervals of volatilities and realized volatilities for stock 600028

和波动率的重现时间间隔分布不同,收益率的重现时间间隔尾分布呈幂律分布形态. 假设收益率的重现时间间隔的尾分布满足幂律分布

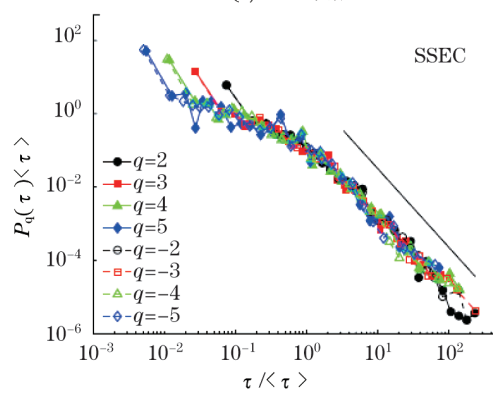
$$P_q(\tau)\langle\tau\rangle = f(\tau/\langle\tau\rangle) = f(x) = cx^{-\delta}, \quad x \geq x_{\min} \quad (18)$$

我们将不同阈值下 q 的标度重现时间间隔 $\tau/\langle\tau\rangle$ 合并,用 Newman^[27] 的极大似然估计方法估计它们共同的拟合分布参数,然后用 KS 拟合优度检验方法检验不同阈值的尾分布是否和它们共同的拟合分布一致. 表 2(见下页)罗列了对上证综指、深证综指和 20 只个股的 p -value 值. 在 1% 的显著水平下,若采用 KS 统计量,上证综指通过拟合优度检验,若同时

采用 1.1 节中的重现时间间隔定义,计算收益率大于阈值 $q>0$ 和小于阈值 $q<0$ 的重现时间间隔. 图 7 给出了上证综指和深证综指在不同阈值下的重现时间间隔分布,可以看出阈值为负和阈值为正分布是重合的,因此分布具有式(14)的标度率行为.



(a) 上证综指



(b) 深证综指

图7 上证综指(SSEC)和深证综指(SZCI)收益率的重现时间间隔分布

Fig.7 Probability distributions of the recurrence intervals of returns for SSEC and SZCI

采用 KS 和 KSW 统计量,上证综指和 16 只个股通过拟合优度检验,说明对于大多数股票,收益率重现时间间隔的尾分布具有标度率,且分布服从幂律分布.

研究还发现阈值 q 和重现时间间隔的均值 $\langle\tau\rangle$ 存在幂律关系

$$\frac{1}{\langle\tau\rangle} = \frac{k}{\beta} |q|^{-\beta} \quad (19)$$

上证综指和深证综指收益率的重现时间间隔均值和阈值关系如图 8(见下页)所示,该结果和同一时期发表在《美国科学院院报》上的一篇文章的结果一致^[34]. 并结合条件重现时间间隔 $\langle\tau|\tau_0\rangle$ 的结果,提出了一个风险估计模型

$$p_{q,\tau_0}^* = \frac{1}{\langle \tau | \tau_0 \rangle} = \frac{1}{\left[1 + \gamma \left(\frac{\tau_0}{\langle \tau \rangle}\right)^{-\mu}\right] \left(\frac{\tau_0}{\langle \tau \rangle}\right)^{\nu}} \frac{k}{\beta} |q|^{-\beta} \quad (20)$$

件的概率. 由式(20)估计得到的条件损失概率 p_{q,τ_0}^* 和实证数据的结果比较符合, 如图 9 所示. 进一步研究表明, 短期记忆性和长期记忆性也存在于收益率的重现时间间隔序列中.

p_{q,τ_0}^* 为条件损失概率, 表示若前一次收益率低于阈值 q 的事件出现在 τ_0 时刻前, 紧接着再发生类似事

表 2 上证综指、深证综指和 20 只个股的 KS 拟合优度检验结果
Tab.2 KS-goodness-of-fit test for the individual stocks in Shanghai and Shenzhen Stock Exchanges

股票代码	$\hat{x}_{\min}$	δ	c	KS	p_{ks}	p_{ksw}
SSEC	3.39	2.20(0.03)	0.29	0.03	0.033	0.000
SZCI	3.33	2.15(0.03)	0.24	0.04	0.010	0.103
600000	4.95	2.78(0.04)	1.30	0.05	0.000	0.058
600009	4.44	2.75(0.04)	1.12	0.02	0.035	0.356
600016	4.86	3.01(0.05)	1.97	0.03	0.016	0.216
600058	5.61	2.95(0.05)	1.87	0.04	0.005	0.098
600060	6.51	3.11(0.07)	2.57	0.03	0.143	0.282
600073	18.36	3.72(0.21)	28.76	0.03	0.968	0.719
600088	4.97	2.94(0.05)	1.73	0.03	0.023	0.163
600098	10.11	3.65(0.12)	14.33	0.03	0.374	0.519
600100	4.90	2.84(0.04)	1.56	0.05	0.000	0.054
600104	5.15	3.04(0.05)	2.25	0.03	0.006	0.013
000001	9.31	3.23(0.09)	4.94	0.04	0.097	0.650
000002	4.89	2.95(0.05)	1.79	0.03	0.013	0.134
000021	5.41	3.26(0.06)	3.49	0.03	0.027	0.170
000027	5.25	3.18(0.05)	2.98	0.03	0.011	0.283
000029	4.72	2.76(0.04)	1.14	0.03	0.010	0.256
000031	11.67	3.90(0.15)	30.43	0.04	0.492	0.713
000039	5.16	3.22(0.06)	2.99	0.03	0.010	0.121
000060	11.28	3.38(0.12)	7.11	0.02	0.963	0.597
000063	5.32	3.29(0.06)	3.71	0.02	0.102	0.460
000066	10.65	3.44(0.10)	9.94	0.03	0.475	0.444

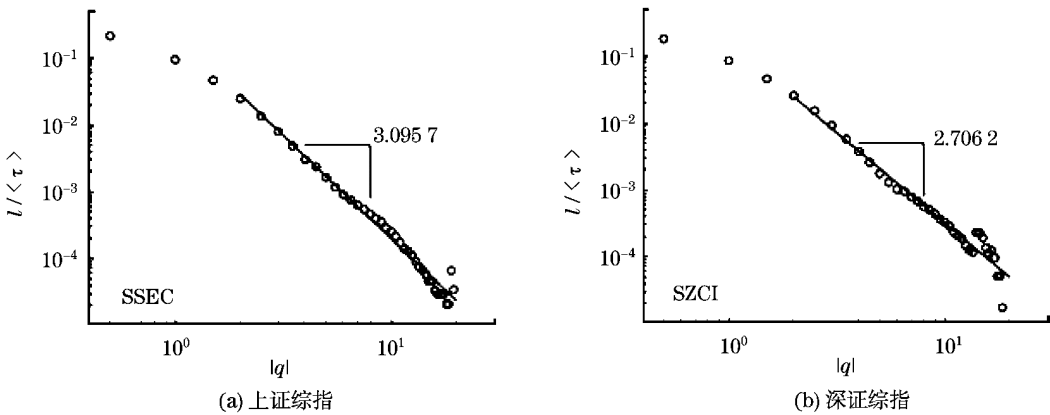


图 8 上证综指和深证综指收益率的重现时间间隔均值和阈值关系
Fig.8 Reciprocal of mean recurrence interval $1/\langle \tau \rangle$ as a function of absolute threshold $|q|$ for SSEC and SZCI

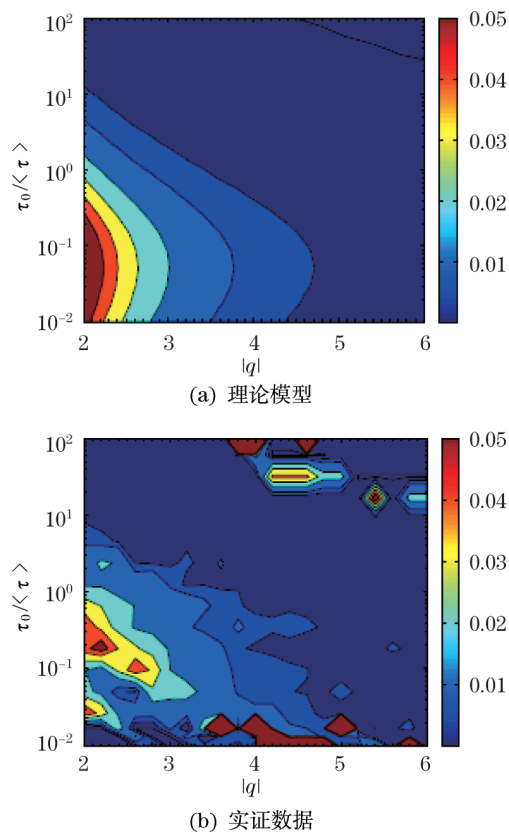


图9 上证综指理论模型和实证数据的条件损失概率的等高图

Fig.9 Contour maps of theoretical loss probability and empirical loss probability

2.4 交易量(trading volume)

最近,重现时间间隔的研究拓展到交易量的研究中^[25-26],并从时间间隔角度考察了大波动和大交易量的关系^[25].采用在上交所和深交所上市的20只个股2000年1月至2009年5月、上证综指和深证综指2003年1月至2009年4月的1分钟数据,分析中国股市交易量的重现时间间隔^[25].研究发现,交易量的重现时间间隔尾分布也服从式(18)的幂律分布,如图10所示,并采用KS拟合优度检验和CvM拟合优度检验验证了尾分布的幂律形式.相关研究还表明,交易量的重现时间间隔序列也具有短期记忆性和长期记忆性.

我们还进一步考察了大交易量时间间隔和大波动时间间隔的同步性,用阈值 q 定义交易量的重现时间间隔, Q 定义波动率的重现时间间隔.图11(见下页)给出的是大波动和大交易量具有相同时间间隔周期的概率,相同波动率阈值 Q 下交易量阈值 q 越大,时间间隔同周期的概率越大,说明大波动通常伴随大交易量同时出现.

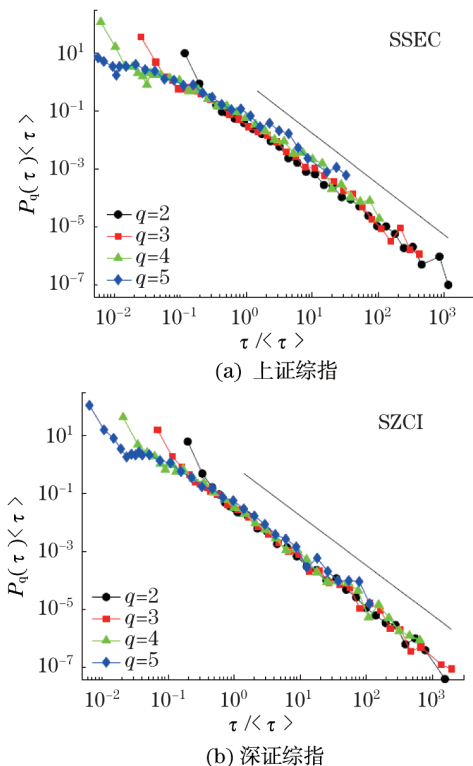


图10 上证综指(SSEC)和深证综指(SZCI)交易量的重现时间间隔分布

Fig.10 Probability distributions of the recurrence intervals of trading volumes for SSE and SZCI

3 微观模型模拟结果

本节将采用委托驱动市场模型^[35],模拟产生股票的价格序列,并运用该价格序列分析收益率的重现时间间隔.首先简单描述一下模型的定义和几个相关参数.委托驱动市场模型是基于交易所内连续委托竞价的交易模式提出的,其构型和实际市场非常接近:市场内有若干个经纪人(agent),每个经纪人根据各自的策略递交委托订单,每个订单包含委托方向(买或者卖)、委托价格和委托量的信息,所有经纪人的委托指令被存储在指令簿(order book)内,系统会根据指令簿内的委托订单信息撮合成交,成交价格即当前股票的价格.

该模型的委托订单包含三方面的信息,首先我们将经纪人的委托量简化为一个基本单位,即每人每次可以买卖一股股票,经纪人委托订单的另外两方面信息由参数 H_s 和 H_x 决定, H_s 用来度量委托方向的记忆性, H_x 用来度量委托价格与指令簿最优价格的相对价格的记忆性.通过调节参数 H_s 和 H_x ,可以观测模拟价格的动力学演化过程,研究不

同参数域下收益率的重现时间间隔.

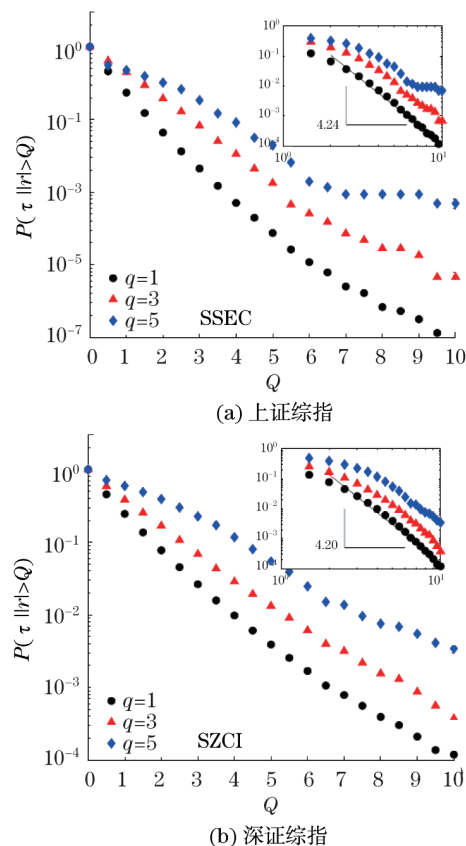


图 11 上证综指 (SSEC) 和深证综指 (SZCI) 大波动和大交易量的时间间隔关系

Fig. 11 Probability of volume intervals conditioned on large volatilities $|r| > Q$ for SSEC and SZCI

图 12 给出了 $H_x = 0.75$ 和 $H_s = 0.5$ 时, 模拟价格的收益率的重现时间间隔分布, 分布的中间部分呈现出幂律分布的形态, 类似于实证数据的幂律分布. 分布的尾巴部分衰落得比较快, 这可能是由于体系的有限尺度效应引起的. 结合上述的两种分布形态, 假设模型收益率的重现时间间隔分布满足以下分布

$$P_q(\tau) \langle \tau \rangle = f(\tau / \langle \tau \rangle) = f(x) = Ax^{-\alpha} e^{-cx^\gamma} \quad (21)$$

采用 Newman^[27] 的极大似然估计方法估计分布参数, 发现幂指数 α 和参数 H_s 没有显著的关系, 和参数 H_x 呈线性关系, 如图 13 所示.

进一步假设幂指数 α 与参数 H_s 和 H_x 存在线性关系

$$\alpha = \beta_0 + \beta_1 H_x + \beta_2 H_s \quad (22)$$

采用最小二乘法得到的线性回归系数为 $\beta_1 = 0.665$, $\beta_2 = -0.017$. 参数 H_x 前的系数在 1% 的显著水平下显著不为零, H_s 前的系数在 1% 的显著水平下不显著不为零, 这和图 13 的结果一致. 分布的

幂指数 α 与参数 H_s 线性相关, 而参数 H_s 决定了收益率 (大于零的单侧收益率) 原始序列的记忆效应, 进一步证实了收益率的重现时间间隔分布和原始序列的记忆性相关.

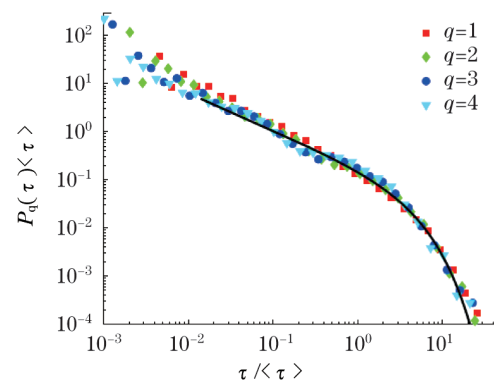


图 12 委托驱动市场模型的重现时间间隔分布

Fig. 12 Probability distribution of the recurrence intervals generated from an order-driven model

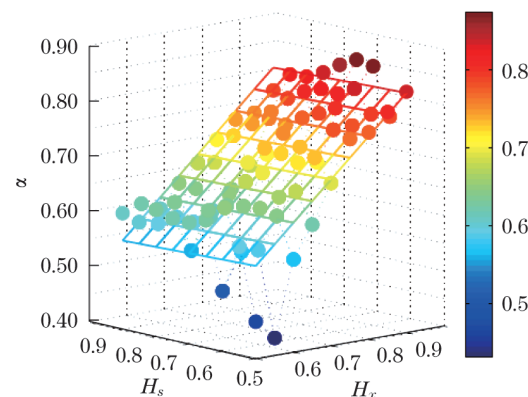


图 13 委托驱动市场模型的重现时间间隔分布的 α 指数与 H_s 和 H_x 参数的关系

Fig. 13 Scaling exponent α of the recurrence intervals generated from an order-driven model as a function of the variables H_s and H_x

4 结 论

本文采用中国股市的高频数据, 对中国金融市场大波动极端事件的重现时间间隔进行研究. 首先介绍了重现时间间隔的概念和基本理论, 并介绍了考察重现时间间隔分布和记忆效应的几种分布检验和关联测度方法, 例如 KS 和 KSW 拟合优度检验、条件概率分布和 DFA 方法. 本文对波动率、已实现波动率、收益率和交易量的重现时间间隔进行了实证分析, 揭示了重现时间间隔分布的非高斯特征, 采用拟合优度检验方法确立了收益率和交易量重现时

间间隔的幂律尾分布形式,进一步验证了重现时间间隔的短期和长期记忆特性,从重现时间间隔分析的角度揭示大波动和大交易量之间的协同关系,并基于收益率重现时间间隔分析,提出了 VaR 风险估计模型.本文还运用委托驱动的微观模型,研究大波动极端事件重现时间间隔的动力学机理,发现重现时间间隔的分布和价格波动序列的记忆性相关,证实记忆效应在大波动极端事件研究中的重要性.

参考文献:

- [1] BOUCHAUD J P. Economics needs a scientific revolution[J]. *Nature*, 2008, 455: 1181.
- [2] LUX T, WESTERHOFF F. Economics crisis[J]. *Nature Physics*, 2009, 5(1): 2-3.
- [3] 张维, 张永杰, 熊熊. 计算实验金融研究[M]. 北京: 科学出版社, 2010.
- [4] 盛昭瀚, 张军, 杜建国. 社会科学计算实验理论与应用[M]. 上海: 上海三联出版社, 2009.
- [5] 盛昭瀚, 张维. 管理科学研究中的计算实验方法[J]. *管理科学学报*, 2011, 14(5): 1-10.
- [6] 周炜星. 金融物理学导论[M]. 上海: 上海财经大学出版社, 2007.
- [7] BUNDE A, EICHNER J F, HAVLIN S, et al. Return intervals of rare events in records with long-term persistence[J]. *Physica A*, 2004, 342(1/2): 308-314.
- [8] BUNDE A, EICHNER J F, KANTELHARDT J W, et al. Long-term memory: A natural mechanism for the clustering of extreme events and anomalous residual times in climate records[J]. *Physical Review Letters*, 2005, 94(4): 048701.
- [9] LIVINA V N, HAVLIN S, BUNDE A. Memory in the occurrence of earthquakes[J]. *Physical Review Letters*, 2005, 95(20): 208501.
- [10] YAMASAKI K, MUCHNIK L, HAVLIN S, et al. Scaling and memory in volatility recurrence interval in financial markets[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(26): 9424-9428.
- [11] WANG F Z, YAMASAKI K, HAVLIN S, et al. Scaling and memory of intraday volatility recurrence intervals in stock markets[J]. *Physical Review E*, 2006, 73(2): 026117.
- [12] OLLA P. Return times for stochastic processes with power-law scaling[J]. *Physical Review E*, 2007, 76(1): 011122.
- [13] SANTHANAM M S, KANTZ H. Return interval distribution of extreme events and long term memory[J]. *Physical Review E*, 2008, 78(5): 051113.
- [14] EICHNER J F, KANTELHARDT J W, BUNDE A, et al. Extreme value statistics in records with long-term persistence[J]. *Physical Review E*, 2006, 73(1): 016130.
- [15] EICHNER J F, KANTELHARDT J W, BUNDE A, et al. Statistics of recurrence intervals in long-term correlated records[J]. *Physical Review E*, 2007, 75(1): 011128.
- [16] JUNG W S, WANG F Z, HAVLIN S, et al. Volatility recurrence intervals analysis of the Japanese market[J]. *European Physical Journal B*, 2008, 62(1): 113-119.
- [17] WANG F Z, YAMASAKI K, HAVLIN S, et al. Indication of multiscaling in the volatility recurrence intervals of stock markets[J]. *Physical Review E*, 2008, 77(1): 016109.
- [18] LEE J W, LEE K E, RIKVOLD P A. Waiting-time distribution for Korean stock-market index KOSPI[J]. *Journal of the Korean Physical Society*, 2006, 48(92): 123-126.
- [19] BOGACHEV M I, BUNDE A. Memory effects in the statistics of interoccurrence times between large returns in financial records[J]. *Physical Review E*, 2008, 78(3): 036114.
- [20] ALYMAN E G, KANTZ H. Recurrence time analysis, long-term correlations, and extreme events[J]. *Physical Review E*, 2005, 71(5): 056106.
- [21] BOGACHEV M I, EICHNER J F, BUNDE A. Effect of nonlinear correlations on the statistics of return intervals in multifractal data sets[J]. *Physical Review Letters*, 2007, 99(24): 240601.
- [22] QIU T, GUO L, CHEN G. Scaling and memory effect in volatility recurrence interval of the Chinese stock market[J]. *Physica A*, 2008, 387(27): 6812-6818.
- [23] REN F, ZHOU W X. Multiscaling behavior in the volatility recurrence intervals of Chinese indices[J]. *Europhysics Letters*, 2008, 84(6): 68001.
- [24] REN F, ZHOU W X. Recurrence interval analysis of high-frequency financial returns and its application to risk estimation[J]. *New Journal of Physics*, 2010, 12(7): 075030.
- [25] REN F, ZHOU W X. Recurrence interval analysis of trading volumes[J]. *Physical Review E*, 2010, 81(6): 066107.
- [26] LI W, WANG F Z, HAVLIN S, et al. Financial factor influence on scaling and memory of trading volume in stock market[J]. *Physical Review E*, 2011. (in press)

(下转第 456 页)