

# 基于模糊信息分配理论的短期股价涨跌模式识别<sup>①</sup>

王新宇<sup>1</sup>, 宋学锋<sup>1</sup>, 吴瑞明<sup>2</sup>

(1. 中国矿业大学管理学院经济管理复杂性研究所, 徐州 221008;

2. 上海交通大学管理学院, 上海 200052)

**摘要:** 基于分形市场假说的股价并不完全反映所有信息的观点, 认为历史股价信息是不完备的群体型模糊信息, 提出了线性信息分配条件下的信息守恒定理, 建立了基于模糊信息分配理论的短期股价涨跌预测的模糊模式识别模型, 通过对上证综合指数日线数据的短期预测, 表明该模型具有能够动态捕捉股价短期分布特征、有效描述股价序列内蕴的短期非线性因果关系, 进而具有较高的股价涨跌识别精度, 并提出了金融市场收益率可能性分布的概念。

**关键词:** 模糊信息分配; 模式识别; 股价涨跌

中图分类号: F830.91

文献标识码: A

文章编号: 1007-9807(2006)02-0061-08

## 0 引言

对股票市场收益趋势、波动性进行预测, 是金融市场投资者和金融领域学者关注的焦点之一。目前国内外的相关研究基本上遵循三种建模路线。首先是基于神经网络、模糊神经网络、小波神经网络、遗传算法等的短期预测模型<sup>[1-6]</sup>; 其次是基于时间序列计量经济模型的预测<sup>[7-11]</sup>, 其中以ARFIMA-GARCH模型族的应用为主。第三种是基于相空间重构的混沌时间序列预测模型, 如用最小邻域线性回归方法预测短期股票走势<sup>[12]</sup>。

大量的实证研究表明金融市场并不遵循有效市场假说 (efficient market hypothesis, EMH), Peters提出了分形市场假说 (fractal market hypothesis, FMH)<sup>[13,14]</sup>, FMH认为市场存在大量的具有不同投资起点的投资者, 并且相互间可以有效转移。价格也不是全部反映所有的信息。EMH假设下的价格体现所有过去的历史信息, 因此价格遵循随机游走过程。但是FMH假设价格只反映部分的信息, 且价格序列中本身也存在一定量噪声的干扰, 即使我们可以得到高频的样本数据, 相对于价格

分布规律内蕴的复杂性而言, 试图通过实际的价格序列精确了解真实信息所引起的价格的分布结构仍是非常困难的。

本文认为股票价格序列是非完备的样本集合, 价格序列信息属于群体型的模糊信息。群体型模糊信息的模糊性本质不在于度量尺度的模糊性, 而是由于信息不全, 现有的信息存在着过渡趋势所引起的, 可以把每个“独立”的价格点, 作为具有一定影响区域但程度有所变化的模糊信息进行处理。长期内股价分布与涨跌模式可能存在结构性改变, 而短期内具有一定的稳定性规律。本文提出基于模糊信息分配方法对短期股价序列的分布规律进行模糊建模、模糊推理和模糊模式识别, 力求揭示短期股价分布内在的非线性规律和复杂性。

## 1 模糊信息分配理论模型

黄崇福系统地论述了模糊信息分配理论<sup>[15]</sup>, 进而将模糊信息分配方法应用到自然灾害风险分析领域<sup>[16]</sup>。王新宇用模糊信息分配方法进行了含有定量

<sup>①</sup> 收稿日期: 2003-05-21; 修订日期: 2004-09-15。

基金项目: 国家自然科学基金资助项目(79970115); 中国矿业大学科技基金资助项目(G200401)。

作者简介: 王新宇(1974-), 男, 江苏新沂人, 博士, 副教授。

和定性指标的采矿复杂系统建模问题求解<sup>[17]</sup>.

### 1.1 一维线性信息分配

记知识样本点的集合为  $W = \{w_1, w_2, \dots, w_m\}$ ,  $V$  为  $W$  的基础论域,  $V_i (i = 1, 2, \dots, n)$  构成对  $V$  的一个划分.  $u_i$  是  $V_i$  的几何中点, 称  $u_i (i = 1, 2, \dots, n)$  为  $V$  上的信息控制点, 记  $U = \{u_1, u_2, \dots, u_n\}$  为  $W$  的离散基础论域. 信息分配的过程就是根据样本点的过渡性质, 对样本点在  $U$  上进行模糊划分的过程. 设  $W, V = [a, b]$  给定时, 取  $u_1 = a, u_n = b$ , 其余的控制点在  $V$  内等距离选取, 常用的一维线性信息分配函数为

$$\mu(u, v) = \begin{cases} 1 - \frac{|u - v|}{\Delta}, & |u - v| < \Delta \\ 0, & \text{其它} \end{cases}$$

( $\Delta$ —控制点步长)

### 1.2 $N$ 维信息分布矩阵 $Q$

设  $W$  为容量为  $m$  的  $N$  元知识样本, 则  $\forall w_j \in W, w_j = (w_{1j}, w_{2j}, \dots, w_{Nj})$ .

记  $T = \{1, 2, \dots, N\}, W_k = \{w_{kj} | j = 1, 2, \dots, m\}, k \in T$ .

显然有:  $W = (W_1, W_2, \dots, W_N), \forall k \in T$ , 设  $W_k$  的基础论域为  $V_k$ , 离散基础论域为  $U_k$ , 则  $W$  的基础论域为  $V = \prod_{k \in T} V_k$ , 相应地离散基础论域为

$$U = \prod_{k \in T} U_k.$$

设  $\mu_{(k)} [\mu_{(k)}(u, v) = \mu_{(k)}(u_k, v_k)]$  是  $W_k$  在  $U_k$  上的信息分配函数,  $\forall v \in V, u \in U$ , 称  $\mu(u, v) = \prod_{k \in T} \mu_{(k)}(u_k, v_k)$  为  $W$  在  $U$  上的信息分配函数.

若取  $\mu_{(k)}$  为线性函数, 则常用的  $N$  维线性分配函数如下

$$\mu(u, v) = \begin{cases} \prod_{k \in T} \left(1 - \frac{|u_k - v_k|}{\Delta_k}\right), & |u_k - v_k| < \Delta_k \\ 0, & \text{其它} \end{cases}$$

设  $U_k$  的离散控制点有  $n_k$  个, 则  $W$  在  $U$  上进行信息分配可得一个  $n_1 \times n_2 \times \dots \times n_N$  维的信息分布矩阵  $Q$ , 记为  $Q = Q(W, U, N) = \{q_{i_1, \dots, i_N} | i_k \in (1, 2, \dots, n_k), k \in T\}$ .  $Q$  的各种内容结构, 就是所要了解的  $W$  的信息结构, 而由  $Q$  生成的模糊关系矩阵  $R$  是其中的一种重要结构. 为了推理

时能方便地基于  $Q$  得  $R$ , 不妨令  $U_k (k = 1, \dots, N - 1)$  是自变量离散论域,  $U_N$  是因变量离散论域,  $X_k = U_k (k \in T_1), T_1 = \{1, \dots, N - 1\}, X = \prod_{k \in T_1} X_k, Y = U_N$ . 记  $Q = Q(X, Y) = \{q(x, y)\}, R = R(x, y) = \{r(x, y)\}$ .

### 1.3 线性信息分配下的信息守恒定理

在模糊信息分配的处理中, 不论是单个样本还是总体, 离散化分配后的信息量应当保持和原始信息量一致, 否则就造成信息损失, 进而模糊推理和识别结果的可靠性与精度会降低. 信息守恒律是这一类方法的一个基本要求, 对多维线性信息分配的情况, 提出下面的定理.

**定理 1** 容量为  $m$  的  $N$  元知识样本点的集合为  $W = \{w_1, w_2, \dots, w_m\}$ , 经过多维线性信息分配得到的信息分布矩阵  $Q = Q(W, U, N) = \{q_{i_1, \dots, i_N} | i_k \in (1, 2, \dots, n_k), k \in T\}$ , 则  $Q(W, U, N)$  的所有多维信息控制点的信息量之和为  $W$  的样本容量  $m$ .

**证明**  $\forall l \in m$  将样本  $w_l \in W$  在  $U = \prod_{k \in T} U_k$  上进行多维线性信息分配. 易知  $w_l$  在  $U_k, \forall k \in T$  上最多使得  $U_k$  中的两个信息控制点获得非零信息量, 记这两个点为  $i_k^1, i_k^2, \sum_{j_k=1}^2 q(i_k^j) = 1$ , 则根据多维线性信息分配函数的定义易知

$$\begin{aligned} & \sum_{j_1=1}^2 \sum_{j_2=1}^2 \dots \sum_{j_N=1}^2 q(i_1^{j_1}, i_2^{j_2}, \dots, i_N^{j_N}) \\ &= \sum_{j_1=1}^2 q(i_1^{j_1}) \sum_{j_2=1}^2 \dots \sum_{j_N=1}^2 q(i_2^{j_2}, \dots, i_N^{j_N}) \\ &= \sum_{j_1=1}^2 q(i_1^{j_1}) \sum_{j_2=1}^2 q(i_2^{j_2}) \sum_{j_3=1}^2 \dots \sum_{j_N=1}^2 q(i_3^{j_3}, \dots, i_N^{j_N}) \end{aligned}$$

依此类推

$$\begin{aligned} & \sum_{j_1=1}^2 \sum_{j_2=1}^2 \dots \sum_{j_N=1}^2 q(i_1^{j_1}, i_2^{j_2}, \dots, i_N^{j_N}) \\ &= \sum_{j_1=1}^2 q(i_1^{j_1}) \sum_{j_2=1}^2 q(i_2^{j_2}) \dots \sum_{j_N=1}^2 q(i_N^{j_N}) = 1 \end{aligned}$$

即  $w_l$  样本形成的多维信息控制点的信息量之和为 1, 于是  $m$  个样本形成的多维信息控制点的累积信息量为  $m$ . 该定理可称为线性信息分配下的信息守恒定理.

### 1.4 多维信息控制点的编码

当指标和控制点的数量增大时,模糊信息分配应用涉及的计算量和复杂程度提高,这在一定程度上限制了该方法的使用.在模糊信息分配方法的编程实现中,必须对多维信息控制点  $i_1 i_2 \dots i_N$  进行编码,以便标记多维信息控制点获得的信息量,该信息量为  $q_{i_1, \dots, i_N} = q(i_1, i_2, \dots, i_N | i_k \in \{1, 2, \dots, n_k\}, k \in N)$ ,在模糊推理中也要用到.令  $n_0 = \max\{n_1, n_2, \dots, n_N\} + 1$ ,我们提出一种采用  $n_0$  进制的  $N$  位数字编码方法,即构造一个  $n_0$  进制的数,具有  $N$  位,第  $k \in T$  位上的数是样本在第  $k$  个指标上的具有非零信息量的一维信息控制点代号  $i_k \in \{1, 2, \dots, n_k\}$ .使用这种方法能够方便地根据指标的个数以及每个指标的信息控制点数的变化进行动态非重复编码,实际程序中将这种  $n_0$  进制数转化为十进制数表示,即

$$f(i_1, i_2, \dots, i_N | i_k \in \{1, 2, \dots, n_k\}, k \in N) = \sum_{k=1}^N i_k (n_0)^{k-1}.$$

### 1.5 基于模糊信息分配的模糊关系矩阵 $R$

#### 1.5.1 基于模糊概念生成的 $R_f$

根据汪培庄提出的因素空间理论<sup>[18]</sup>,当  $X \times Y$  是一般的因素空间时,一个元素对应一个模糊概念是允许的.设  $Y$  是一个以模糊概念构成的状态空间,  $\forall y \in Y$  是一个模糊概念.  $X$  可以视为  $Y$  的基础论域,可以由  $Q$  得出  $X$  关于  $Y$  的隶属度,形成模糊关系矩阵  $R_f$

$$R_f = \{q(x, y) / s(y)\} \\ s(y) = \max_{x \in X} \{q(x, y)\}$$

#### 1.5.2 基于条件落影公式生成的 $R_s$

利用集值统计和落影理论中的条件落影公式生成  $R_s$ ,在  $W$  生成  $Q$  的过程中,经过信息分配,  $W$  原本在  $N$  维空间中的每一点都变成了一个模糊集合.对各点分配的信息量值就是该模糊集的隶属度值,这些模糊集合叠加在一起形成  $Q$ .

因此可由  $Q$  来求  $W$  在  $X \times Y$  上的联合落影  $\mu_{\xi, \eta}(x, y)$  和关于  $X$  的边际落影  $\mu_{\xi}(x)$

$$S = \int_X \int_Y q(x, y) dx dy \\ \mu_{\xi, \eta}(x, y) = q(x, y) / S \\ \mu_{\xi}(x) = 1/S \int_Y q(x, y) dy$$

从而得条件落影

$$\mu_{\eta/\xi}(y | x) = \mu_{\xi, \eta}(x, y) / \mu_{\xi}(x) = \frac{q(x, y)}{\int_Y q(x, y) dy}$$

由于  $X \times Y$  已经离散化了,所以条件落影表达式为

$$\mu_{\eta/\xi}(y_j | x_i) = \frac{q_{ij}}{\sum_{j=1}^L q_{ij}}$$

则模糊关系矩阵  $R_s = \{r_{ij}\} = \{\mu_{\eta/\xi}(y_j | x_i)\}$

### 1.6 由 $R$ 进行模糊近似推理

模糊推理的过程本质上是寻求从  $X$  上的模糊集合  $\underline{A}$  到  $Y$  上的模糊集合  $\underline{B}$  的模糊变换,利用已经通过信息分配方法建立的模糊关系矩阵  $R$  导出一个模糊变换来实现该变换,即

$$\underline{B} = \underline{A} \times R$$

在  $R$  已经确定的情况下,模糊近似推理主要包含下面两个内容.

#### 1.6.1 $\underline{A}$ 的确定

设  $w$  为待识别样本,将  $\forall w \in W$  在由输入变量构成的离散基础论域即笛卡尔乘积空间  $X = \prod_{k \in T_1} X_k = \prod_{k \in T_1} U_k$  上进行信息分配,所得到的信息控制点的信息分配量  $q_i$  构成  $X$  上的模糊集合

$$\underline{A} = (q_1, q_2, \dots, q_m) \\ (m \text{ 为 } X \text{ 上的信息控制点数目})$$

$\underline{A}$  的含义是样本  $w$  隶属于各个信息控制点的程度.

#### 1.6.2 模糊运算算子的选择

##### 1) $R_f$ 上的推理

##### ① $R_f$ 上的 max-min 推理

$$\forall y \in Y, \underline{B}(y) = \sup_{x \in X} \{ \underline{A}(x) \wedge \underline{R}(x, y) \}$$

##### ② $R_f$ 上的格式贴程度推理

$$\forall y \in Y, \underline{B}(y) = \left[ \sup_{x \in X} \{ \underline{A}(x) \wedge \underline{R}(x, y) \} \right] \wedge \left[ 1 - \inf_{x \in X} \{ \underline{A}(x) \vee \underline{R}(x, y) \} \right]$$

##### ③ $R_f$ 上的加权求和推理

$$\forall y \in Y, \underline{B}(y) = \min \left\{ 1, \sum_{x=1}^m \underline{A}(x) \underline{R}(x, y) \right\}$$

##### 2) $R_s$ 上的推理

由条件落影公式得到  $R_s$ ,需要用全落影公式

进行推理

$$\forall y \in Y, \tilde{B}(y) = \frac{\int_X \tilde{A}(x) R(x, y) dx}{\int_X \tilde{A}(x) dx}$$

由于  $X \times Y$  是离散化了的, 实际的推理公式为

$$\forall y_j \in Y, \tilde{B}(y_j) = \frac{\sum_{i=1}^m \tilde{A}(x_i) R(x_i, y_j)}{\sum_{i=1}^m \tilde{A}(x_i)}$$

( $m$  为  $X$  中信息控制点数目)

## 2 实证分析

本文采用“钱龙系统”提供的上海证券交易所的上证综合指数日收盘价格, 使用对数收益率  $x_i = \ln(P_i) - \ln(P_{i-1})$ ,  $P_i$  为收盘价格, 实证采用 Matlab 语言编程计算.

### 2.1 数据的基本统计特征描述

数据的基本统计特征描述见表 1, 可以看出收益分布是有偏和尖峰的, 对数收益率曲线见图 1.

表 1 上海股市收益率的基本统计特征

Table 1 Descriptive statistics of returns for Shanghai stock market

指数	样本区间	样本数	均值	最大值	最小值	标准差	偏度 Skewness	峰度 Kurtosis	Jarque-Bera 统计量
上证综合	2002/06/25 —2002/12/31	129	-0.001 773	0.027 734	-0.027 216	0.009 257	-0.331 891	3.807 913	5.876 654

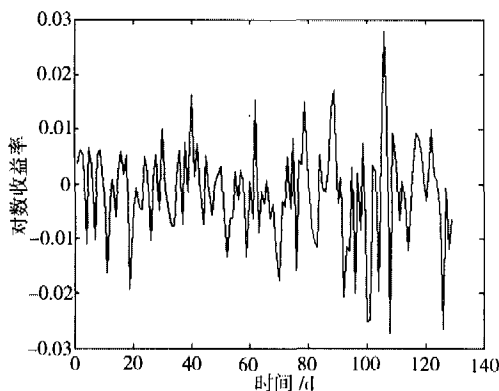


图 1 对数收益率

Fig. 1 Plot of logarithmic returns

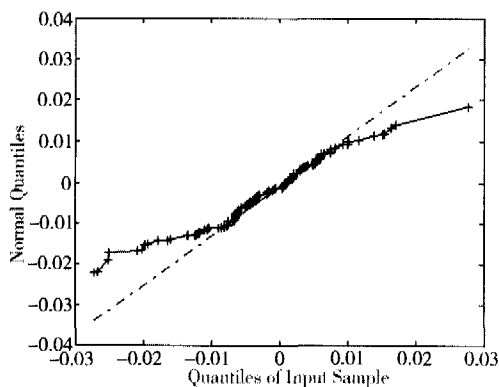


图 2 经验分布相对正态分布的 Q-Q 图

Fig. 2 Q-Q plot for sample distribution against normal distribution

上证综合指数日收益经验分布相对正态分布的 Q-Q 图(典型的非直线形状) 见图 2, 它清楚

显示经验分布偏离正态分布.

### 2.2 数据预处理

模糊信息分配方法应用首先要确定输入输出指标的个数及每个指标的信息控制点. 中国股票市场从 1996 年 12 月 16 日开始实行涨跌  $\pm 10\%$  停板制度, 输入和输出指标的信息控制点的最大区间是  $(\log 0.9, \log 1.1)$ . 指标控制点的选取应首先取决于领域的专家知识, 即用指标的关键点(如工程上指标的临界点或分级点), 但在金融市场上收益率是连续的, 很难确定“关键点”. 为了简化计算, 本文用等间距的信息控制点, 也可以采取非等间距的信息控制点, 比如可以按照极端收益率区域较少, 而均值收益率区域较多的方式选取. 当样本容量相对较大时, 样本空间的信息密度较大, 可以适度增多控制点的个数; 反之亦然. 控制点的个数也不宜太少, 否则不足以描述复杂的分布规律, 同时输出指标识别结果的分辨率不高. 输入和输出指标的信息控制点数可以不同.

多维样本用一维序列滞后数据构造, 多维样本数据有交叉部分, 以在一定的样本容量内获得更多的多维样本, 即更多的股票价格涨跌模式.

### 2.3 模糊推理与识别

考虑了不同的输入指标数以及输入和输出控制点的多种组合, 用 10 个样本检验模型效果, 金融投资者对股价涨跌方向的关心程度比股价数值要大, 我们从涨跌方向是否一致来检验模型的有

效性,见表 2.可以看出,输入指标数对识别精度影响最大,用滞后 4 d 和 7 d 的数据作为输入指标时,其平均识别精度相对较低;滞后 6 d 的数据作为输入,则仅当用 9 个输入控制点在  $R_s$  上的推理效果良好.用 5 d(一周)滞后数据作为输入,平均识别效果最好,部分原因可能是由于金融市场普遍存在的星期效应.另外,市场对新信息的接受理解需要一个过程,对信息的反应有一个时滞过程,时滞过程的长短取决于股票市场的发育程度,张思奇等认为上海股票市场不超过 5 d<sup>[6]</sup>,本文也得出类似的结论.

用滞后 5 d 对数收益作为输入指标,构成 124 个多维样本,前 114 个样本用于建模,其余用于检

验模型的精度.首先用建模样本在输入和输出指标构成的多维信息控制点进行信息分配,将各多维信息控制点获得的信息量累加即生成原始信息布矩阵,由于生成的原始信息分配矩阵的规模很大(检验发现其所有多维信息控制点的信息量之和为 114,等于建模样本容量,与信息守恒定理结论一致),在这里仅给出一部分,见表 3.由  $Q$  生成  $R$  就可以进行模糊推理.对检验样本采用  $R_s$  上的全落影公式推理,结果见表 4,识别准确率达到 80%;而  $R_f$  上的推理对检验样本的识别准确率则达到 90%.不同模糊算子对检验样本的识别精度是有差异的,实际应用中选择目前精度最好的算子用于新样本的识别.

表 2 股价涨跌方向识别精度比较

Table 2 Ratios of correct recognition to fluctuation direction in models

输入指标的个数	各输入指标的信息控制点数	输出指标的信息控制点数	$R_s$ 上的识别结果			$R_f$ 上的识别结果		
			建模样本 (%)	检验样本 (%)	总样本 (%)	建模样本 (%)	检验样本 (%)	总样本 (%)
4	6	5	57	30	55	64	40	62
		6	77	70	77	64	40	62
		7	63	50	62	66	50	65
	7	6	83	40	79	74	40	71
		7	64	50	63	64	30	62
		8	85	50	82	72	40	70
	8	7	73	70	73	76	80	76
		8	89	60	86	71	80	72
		9	80	60	78	78	90	79
5	7	6	94	70	92	84	40	81
		7	74	60	73	79	50	77
		8	94	60	91	79	50	77
	8	6	95	80	94	87	90	87
		7	79	90	80	81	90	81
		8	96	80	95	89	90	89
	9	9	87	80	86	84	90	85
		8	99	60	96	95	70	93
		9	89	70	88	89	70	87
6	7	10	99	60	96	97	60	94
		6	99	60	96	74	50	72
		7	78	40	75	85	50	82
	8	8	98	70	96	81	50	79
		7	81	50	78	86	20	80
		8	98	70	96	96	20	90
	9	9	89	60	86	86	30	81
		8	100	90	99	89	30	85
		9	89	90	89	90	30	85
7	10	100	90	99	92	30	87	
	8	99	40	94	90	40	86	
	9	89	40	85	91	40	87	
7	10	100	40	95	84	40	80	

表 3 原始信息分布矩阵(部分)  
Table 3 Original information distribution matrix (partial)

输入指标多维 信息控制点编码	输出指标信息控制点编码							
	1	2	3	4	5	6	7	8
56645	0.023	0.099 6	0	0.000 1	0.001 3	0	0	0
56646	0	0	0	0.000 1	0.000 6	0	0	0
55354	0	0.005 9	0.106 9	0.053	0.007	0.109 7	0.102 2	0
55355	0	0.000 3	0.370 1	0.577	0.105 5	0.016 9	0.019 7	0
55364	0	0	0.022 4	0.039 7	0	0.033	0.039 7	0
55365	0	0	0.141 2	0.079 7	0.013 8	0.012 5	0.015	0
55454	0	0.009	0.117 1	0.104 3	0.020 6	0.044 8	0.053 8	0
55455	0	0.000 4	0.039	0.133 8	0.066 8	0.016 9	0.020 3	0
55464	0	0	0.040 8	0.196 9	0	0.034 1	0.040 9	0
55465	0	0	0.023 5	0.082 2	0.002 5	0.012 9	0.015 5	0
56354	0	0	0.006	0.002 9	0	0.014 5	0.017 4	0
56355	0	0	0.055 2	0.010 5	0	0.005 5	0.006 6	0
56364	0	0	0.003 5	0.005 9	0	0.011	0.013 2	0
56365	0	0	0.024	0.006 3	0	0.004 2	0.005	0
56454	0	0	0.012 3	0.021 7	0	0.014 9	0.017 9	0
56455	0	0	0.007 5	0.008 3	0	0.005 6	0.006 8	0
56464	0	0	0.013 7	0.040 6	0	0.011 4	0.013 6	0
56465	0	0	0.006 2	0.015 1	0	0.004 3	0.005 2	0
65354	0	0.011 7	0.137 2	0.027	0.001 3	0	0	0
65355	0	0.000 5	0.123 2	0.164 2	0.039	0	0	0

表 4 检验样本在  $R_i$  上的模式识别结果

Table 4 Recognition result for test samples by  $R_i$

检验样 本编号	输出指标的各个信息控制点的隶属度								对数收益率		涨跌方向 是否一致
	1	2	3	4	5	6	7	8	实际值	识别值	
115	0	1	0.046	0.331 5	0.602 3	0.136 8	0.002 2	0	-0.003 1	-0.019 4	是
116	0.195 8	0.894 5	1	0.755 6	0.154 1	0.020 7	0.001 5	0	0.001 3	-0.011 5	否
117	0.220 9	1	0.47	0.620 4	0.850 5	0.088 5	0.001 7	0	0.010 1	-0.019 4	否
118	0.152 7	0.692 4	0.199 3	0.747 5	0.380 8	0.573	1	0	0.001 6	0.019 9	是
119	0	0.191 4	0.455 4	0.437 7	1	0.213 2	0.007 3	0	0.000 4	0.004 2	是
120	0	1	0.264 9	0.449 1	0.307 4	0.024 7	0.001	0	-0.006 5	-0.019 4	是
121	0	0.040 6	1	0.73	0.183 8	0.041 6	0.042 6	0	-0.026 6	-0.011 5	是
122	0	0.015 4	1	0.732 9	0.16	0.047 9	0.04	0	-0.000 9	-0.011 5	是
123	0	0.016 5	1	0.766 5	0.132 1	0.040 3	0.043 3	0	-0.012 1	-0.011 5	是
124	0	0.016 7	1	0.757 7	0.199 8	0.083 2	0.073 6	0	-0.006 4	-0.011 5	是

2.4 股价对数收益率的可能性分布

从模糊信息分配理论和应用中可以看到,所得到的模式识别结果是一个模糊集合,应用在股票市场中,就是一个关于股票价格对数收益的模糊集合  $B$ ,此模糊集合的隶属函数实际上提供了股票收益变化的可能性分布.相比之下,一般的神经网络、混沌时间序列模型无法得到这方面的信息,而时间序列计量模型得到的是基于概率分布的置信区间.

定义 1  $Y = \{y_i, i = 1, 2, \dots, n\}$  为股票价格对数收益率的离散论域,  $\pi(y_i), i = 1, 2, \dots, n$  是事件  $y_i$  发生的可能性,其中  $\pi(y_i) = \underline{B}(y_i), i = 1, 2, \dots, n$ , 则称为  $\pi(Y)$  是股价收益的可能性分布函数.

$B$  的  $\lambda$  水平截集为  $B_\lambda = \{y | y \in Y, \underline{B}(y) \geq \lambda\}$ , 则  $B_\lambda$  是置信水平为  $\lambda$  时股价对数收益率的可能性区间, 可以作为投资风险的一种参考依据. 例如, 图 3 是根据对检验样本 119 和 123 的识别结

果得到的对数收益率可能性分布。

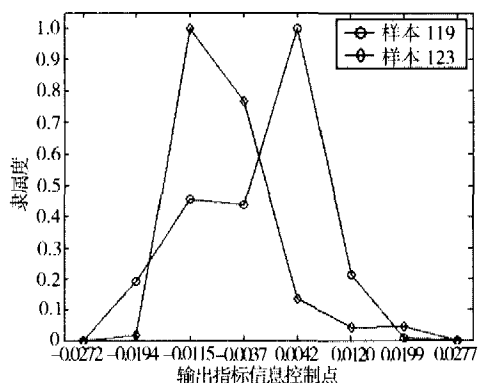


图 3 检验样本 119 和 123 的对数收益率可能性分布

Fig. 3 Possibility distributions of logarithmic returns for Sample 119 and 123

### 3 主要结论

在分形市场假说基础上,提出可以认为股价序列信息是一种群体型模糊信息,适合采用模糊信息优化处理技术来解决.提出了在多维线性信

息分配函数下的信息守恒定理,首次采用模糊模式识别的思想,运用模糊信息分配理论解决了传统的金融市场时间序列建模问题,它既是一种金融市场时间序列分析的新工具,又发展了模糊信息分配理论的应用范围.

由于股价分布与涨跌规律非常复杂,传统的时间序列计量经济模型均要求收益率服从特定的概率分布,但是对实际收益率分布具有的尖峰胖尾、非正态性等特征仍然没有彻底的解决.模糊信息分配模型不需要设定收益率的概率分布形式,把样本信息离散化分配与叠加,通过样本数据的原始信息分布矩阵来揭示样本总体信息的真实精细结构,增强了对总体的认识.由原始信息分配矩阵生成的模糊关系矩阵实际上从样本中学习集成了股价分布规律的知识,并且这个知识库可以随着新股价信息的到达不断更新.实证分析表明,该模型对股价涨跌方向的识别精度是较高的,同时能够提供未来收益率的可能性分布,用以指导投资决策.进一步的研究工作是在输入指标中加入交易量序列,以期提高模型的有效性.

### 参考文献:

- [1]陈兴,孟卫东,严太华.基于 T-S 模型的模糊神经网络在股市预测中的应用[J].系统工程理论与实践,2001,21(2):66—72.  
Chen Xing, Meng Wei-dong, Yan Tai-hua. Forecasting stock market quotations via fuzzy neural network based on T-S model[J]. System Engineering Theory and Practice, 2001, 21(2): 66—72. (in Chinese)
- [2]解光军,庄镇泉.差分 RBF 神经网络的预测算法及其应用[J].信息与控制,2000,29(5):421—424.  
Xie Guang-jun, Zhuang Zhen-quan. A predictive algorithm based on difference radial basis function neural networks mode for time series and its application[J]. Information and Control, 2000, 29(5): 421—424. (in Chinese)
- [3]Joao C. Teixeira, Antonio J. Rodrigues. An applied study on recursive estimation methods, neural networks and forecasting[J]. European Journal of Operational Research, 1997, 101: 406—417.
- [4]An-Sing Chen, Mark T. Leung, Hazem Daouk, Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index[J]. Computers & Operations Research, 2003, 30: 901—923.
- [5]William Leigh, Russell Purvis, James M. Ragusa. Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support[J]. Decision Support Systems, 2002, 32: 361—377.
- [6]程乾生,王守章,武连文.基于遗传算法的多阶马氏链组合预测方法[J].管理科学学报,1998,1(6):26—33.  
Cheng Qian-sheng, Wang Shou-zhang, Wu Lian-wen. The combined prediction method based on genetic algorithm and multiple-order markov chains[J]. Journal of Management Sciences in China, 1998, 1(6): 26—33. (in Chinese)
- [7]张思奇,马刚,冉华.股票市场风险、收益与市场效率—ARMA-ARCH-M 模型[J].世界经济,2000,(5):19—28.  
Zhang Si-qi, Ma Gang, Ran Hua. Risk, return and efficiency in stock markets: ARMA-ARCH-M model[J]. World Economy, 2000, (5): 19—28. (in Chinese)
- [8]樊智,张世英.多元 GARCH 建模及其在中国股市分析中的应用[J].管理科学学报,2003,6(2):68—73.

- Fan Zhi, Zhang Shi-ying. Multivariate GARCH modeling and its application in volatility analysis of Chinese stock markets[J]. Journal of Management Sciences in China, 2003, 6 (2): 68—73. (in Chinese)
- [9]魏巍贤, 周晓明. 中国股票市场波动的非线性 GARCH 预测模型[J]. 预测, 1999, 18(5): 47—49.  
Wei Wei-xian, Zhou Xiao-ming. Forecasting the volatility of Chinese stock market with GARCH[J]. Forecasting, 1999, 18(5): 47—49. (in Chinese)
- [10]Philip Hans Franses, Hendrik Ghijsels. Additive outliers, GARCH and forecasting volatility[J]. International Journal of Forecasting, 1999, 15: 1—9.
- [11]Peter Verhoeven, Berndt Pilgram, Michael McAleer, Alistair Mees. Non-linear modelling and forecasting of S&P 500 volatility[J]. Mathematics and Computers in Simulation, 2002, 59: 233—241.
- [12]Fernando Fernandez-Rodríguez, Simon Sosvilla-Rivero, Mara Dolores Garcia-Artiles. Dancing with bulls and bears: Nearest-neighbour forecasts for the Nikkei index[J]. Japan and the World Economy, 1999, 11: 395—413.
- [13]Edgar E. Peters. *Appling Chaos Theory to Investment and Economics*[M]. New York: John Wiley & Sons, 1994.
- [14]Edgar E. Peters. *Chaos and Order in the Capital Markets*[M]. New York: John Wiley & Sons, 1996.
- [15]黄崇福, 王家鼎. 模糊信息优化处理技术及其应用[M]. 北京: 北京航空航天大学出版社, 1995.  
Huang Chong-fu, Wang Jia-ding. *Fuzzy Information Optimization Technology and its Applications*[M]. Beijing: Peking University of Aeronautics and Astronautics Press, 1995. (in Chinese)
- [16]黄崇福. 自然灾害风险分析的基本原理[J]. 自然灾害学报, 1999, 8(2): 21—30.  
Huang Chong-fu. Basic principles of risk analysis of natural disasters[J]. Journal of Natural Disasters, 1999, 8(2): 21—30. (in Chinese)
- [17]王新宇, 俞书伟, 王洪欣. 基于模糊信息分配方法的综采工艺选择模型[J]. 系统工程理论与实践, 2000, 20(7): 139—144.  
Wang Xin-yu, Yu Shu-wei, Wang Hong-xin. A model based on the theory of fuzzy information distribution for selecting coal mining method[J]. System Engineering Theory and Practice, 2000, 20(7): 139—144. (in Chinese)
- [18]汪培庄. 模糊集与随机集落影[M]. 北京: 北京师范大学出版社, 1985. 42—69.  
Wang Pei-zhuang. *Fuzzy Set and Falling Shadow of Random Set*[M]. Beijing: Peking Normal University Press, 1985. 42—69. (in Chinese)

## Pattern recognition model of short-term share price fluctuation based on fuzzy information distribution theory

WANG Xin-yu<sup>1</sup>, SONG Xue-feng<sup>1</sup>, WU Rui-ming<sup>2</sup>

1. Economic and Managerial Complexity Institute, School of Management, China University of Mining and Technology, Xuzhou 221008, China;
2. Management School, Shanghai Jiaotong University, Shanghai 200052, China

**Abstract:** According to the fractal market hypothesis, the current share prices do not absorb all the historical market information, we think that the price series information is a sort of grouped incomplete fuzzy information. The theorem that the information volume is a constant under the linear information distribution function is proved. We set up a pattern recognition model, based on the fuzzy information distribution theory, to predict the direction of fluctuation of Shanghai stock market composite index day time series. By empirical analysis it is found that this model is very effective and can also give the possible distribution of logarithmic stock return. Finally the framework of decision support system for pattern recognition of short-term share price fluctuation is put forth.

**Key words:** fuzzy information distribution theory; pattern recognition; share price fluctuation