

# 证券市场信息熵技术指标构建与应用

王保山, 邵明星, 丁爱亮, 徐炳振\*

(宁波大学 理学院, 浙江 宁波 315211)

**摘要:** 选取上证每个交易日的日线数据, 并对其进行统计分析. 首先利用小波降噪技术, 得出信息熵技术指标, 然后适当选取信息熵阈值进行分析, 得出价格走势以及市场特点对应的信息熵指标, 进而指出其对长线投资信息熵技术指标的应用具有很好的参考价值.

**关键词:** 信息熵; 小波分析; 熵阈值

中图分类号: O411.1

文献标识码: A

文章编号: 1001-5132 (2012) 03-0079-04

信息是个很抽象很宽泛的概念, 很难用一个简单的定义将其完全准确把握<sup>[1]</sup>. 平日我们所指的信息量的大小, 一般很难得到一个确定的量化数值; 而在证券市场中, 股民大多数是利用自己所掌握的信息来判断何种证券值得持有、持有的时间要多久、何时进场何时出场? 然而, 股民所掌握的信息到底是个什么样子? 信息量越大越好吗? 怎样才能判断这种信息的价值呢? 然而, 一个概率分布可以被定义成一个“熵”的量, 它具有许多特性符合度量信息的直观要求, 因此, 笔者就以上问题做出量化解答, 并就此构建信息熵的技术指标.

笔者以上证指数 000001 的交易时间、开盘、最高、最低、收盘、成交量、成交额、涨数、跌数、持仓量、分笔、开盘量、开盘额数据为信息构建信息熵指标, 选取数据为 1991 年 5 月 13 日至 2012 年 3 月 22 日这 20 年的日线数据; 之所以没有选取 1991 年 5 月 13 日之前的数据, 是因为在此时间之前的数据有些是不完整且不稳定的, 不能体现出信息熵计算的稳定性和可信度.

## 1 信息熵

1948 年, Shannon 借鉴了热力学的概念, 将信息中排除了冗余后的平均信息量称为“信息熵”<sup>[2]</sup>, 并给出了计算信息熵的数学表达式. 对于 1 个随机

的离散变量, 可能有  $n$  个取值, 每个取值所对应的概率分别为  $p_1, p_2, \dots, p_i, \dots, p_n$ , 则其信息熵可表述为下式:

$$H = -\sum_{i=1}^n p_i \log p_i. \quad (1)$$

对于变量的概率值而言, 其值越大, 其不确定性越大, 信息熵也越大, 意味着想要更好地确定其具体情况, 所需要的信息量就越大, 计算所得的信息熵也就越大. 而当  $p_1 = 1, p_2 = p_i = \dots = p_n = 0$  时, 这个随机变量就被确定, 其信息熵最小为 0.

然而, 证券市场的价格走势是不断变化的, 体现的信息熵值也随之不断变化, 因此, 我们只凭大盘 K 线的走势是很难判断下一个周期价格是涨还是跌<sup>[3]</sup>. 所以, 提取出所需要的信息, 并计算得到 1 个适合价格分析的指标就尤为重要, 而此时的信息熵指标就能很好地表现出股市价格走势和市场特点.

## 2 指标模型建立过程<sup>[4]</sup>

(1) 选取数据源. 将数据源中的每 *Circl* 个交易日作为 1 个时间周期单位(文中 *Circl* 为 15), 共分为  $m$  组数据(文中  $m$  为 326), 将每列数据(开盘价、最高价等)作为 1 个小组数据计算信息熵值, 表 1 为数据的分组列表示意形式, 因原数据众多, 因

收稿日期: 2012-03-31.

宁波大学学报(理工版) 网址: <http://3xb.nbu.edu.cn>

基金项目: 国家自然科学基金(10774080).

第一作者: 王保山(1981—), 男, 山东临沂人, 在读硕士研究生, 主要研究方向: 金融物理与复杂系统动力学. E-mail: baoshan2007@163.com

\*通讯作者: 徐炳振(1959—), 男, 山东德州人, 博士/教授, 主要研究方向: 金融物理与复杂系统动力学. E-mail: xubingzhen@nbu.edu.cn

表1 数据源分组列表

| 交易时间       | 开盘价/元  | 最高价/元  | 最低价/元  | 收盘价/元  | 成交量/手 | 成交额/元  | ... |
|------------|--------|--------|--------|--------|-------|--------|-----|
| 1991-05-13 | 107.88 | 107.88 | 107.74 | 107.84 | 235   | 182000 | ... |
| ...        | ...    | ...    | ...    | ...    | ...   | ...    | ... |

此只列示选取的第1个数据格式,其余格式均相同。

(2) 选取每个小组数据第2天数减前1天数,得到相应差值,计算公式如下:

$$ChaZhi = data(i+1) - data(i), \quad (2)$$

其中,  $data(i)$ ,  $data(i+1)$  分别表示前1天和第2天数数据。

(3) 根据计算之后的  $ChaZhi$ , 统计出大于、小于和等于零的  $ChaZhi$  个数  $n(ChaZhi >, <, = 0)$  各为多少, 计算其值所占的比重(即为概率), 计算公式如下:

$$p(ChaZhi >, <, = 0) = \frac{n(ChaZhi >, <, = 0)}{Circel}. \quad (3)$$

(4) 在信息熵值计算中, 由于  $\log(0)$  是无意义的, 所以,  $p(ChaZhi >, <, = 0) = 0$  时, 将其值改为  $p(ChaZhi >, <, = 0) = 10^{-10}$ , 可计算  $(10^{-10}) \log(10^{-10}) \propto 0$ , 所以不会影响熵值计算.  $p(ChaZhi >, <, = 0)$  可简写为  $p(C >, <, = 0)$ , 信息熵计算公式如下:

$$H_m = -[(p(C > 0)) \log(p(C > 0)) + (p(C < 0)) \log(p(C < 0)) + (p(C = 0)) \log(p(C = 0))].$$

(5) 计算每个周期  $Circel$  交易日的总信息熵值的公式为:

$$H_{mi} = \sum_{m=1}^{Circel} H_m.$$

(6) 计算  $m$  组数据总信息熵值的公式为:

$$H = \sum_{i=1}^m H_{mi}. \quad (4)$$

(7) 由于计算得到的信息熵值序列是一个噪音相当大的数值序列, 因此, 笔者采用小波分析降噪的方法将其主要趋势序列进行提取<sup>[5-6]</sup>, 其信息熵原始信号与分层阈值降噪信号对比如图1所示。

(8) 将上证1991年5月13日开始至2012年3月22日的20年交易日日线数据每隔  $Circel$  个收盘价格取1个值, 得到  $m$  个价格时间序列。

(9) 将价格时间序列和信息熵值均作归一化处理<sup>[7]</sup>, 为更加清晰地说明图例, 将每个归一化的信息熵值增加1, 得到调整后的信息熵值序列, 而

价格归一化处理后的价格序列对应如图2所示。

(10) 选取调整后的熵值序列中适当的信息熵

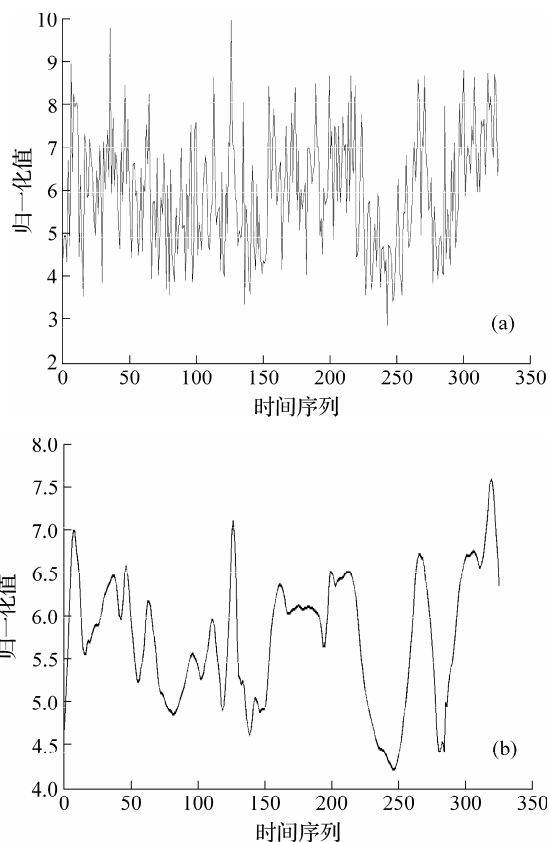


图1 信息熵原始信号(a)与分层阈值降噪信号(b)的对比

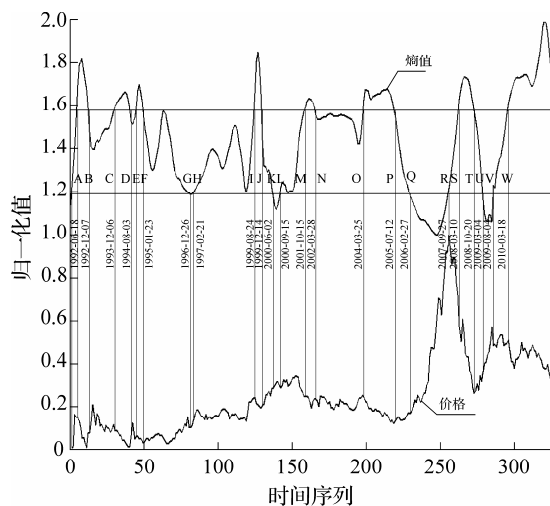


图2 调整信息熵选取特定值与价格时间序列对应

表 2 时间与价格对应表

| 时间         | 价格/元      | 时间         | 价格/元      | 时间         | 价格/元      | 时间         | 价格/元      |
|------------|-----------|------------|-----------|------------|-----------|------------|-----------|
| 1992-06-18 | 1168.4300 | 1996-12-26 | 924.2400  | 2001-10-15 | 1687.2800 | 2008-03-10 | 4146.3000 |
| 1992-12-07 | 793.2500  | 1997-02-21 | 1017.1300 | 2002-03-28 | 1650.6700 | 2008-10-20 | 1974.0100 |
| 1993-12-06 | 995.7200  | 1999-08-24 | 1648.4700 | 2004-03-25 | 1736.4800 | 2009-03-04 | 2198.1100 |
| 1994-08-03 | 522.9800  | 1999-12-14 | 1422.2600 | 2005-07-12 | 1046.1600 | 2009-08-04 | 3471.4400 |
| 1994-10-07 | 738.0300  | 2000-06-02 | 1916.2500 | 2006-02-27 | 1297.1900 | 2010-03-18 | 3046.0900 |
| 1995-01-23 | 553.4400  | 2000-09-15 | 1978.4300 | 2007-09-27 | 5409.4000 |            |           |

阈值<sup>[8-9]</sup>,  $H_a=1.19, H_b=1.58$  (图 2), 然后选取其值对应的时间和收盘价格 (表 2)。

3 实证分析

对图 2 进行分析, 可得到如下信息:

(1) 熵值总是在振荡中, 其变化的相对幅度明显高于价格的相对变化幅度。

(2) 超出  $[H_a=1.19, H_b=1.58]$  范围之外的收盘价走势可以分为两种情况, (1)  $H < H_a=1.19$  价格均为上涨; (2)  $H > H_b=1.58$  价格均为下跌。

(3) 熊市的明显的特点是熵值整体比较高<sup>[2]</sup>, 因此, 利用此点可以判断股市是否处于熊市。熵值在阈值  $H_b=1.58$  (图 2 中水平线) 上方振荡, 则表明股市整体格局较乱, 处于明显的熊市中; 只有当熵值充分下穿  $H_b=1.58$  线, 才可以视为熊市结束的信号 (图 2 中 2010 年 03 月 18 日对应的直线)。而目前 (数据截止日为 2012 年 3 月 22 日) 的熵值仍在阈值上方振荡, 表明熊市没有结束。

(4) 当信息熵值曲线上穿  $H_b=1.58$  时, 价格将要下跌, 下穿  $H_b$  时, 价格将要上涨。

(5) 当信息熵值曲线下穿和上穿  $H_a=1.19$  时, 要注意其信息熵值变化量的大小。如果熵值变化大, 那么, 熵值曲线下穿  $H_a=1.19$  时, 价格将要上涨, 上穿  $H_a=1.19$  时, 价格将要下跌; 熵值变化小, 则不一定有价格趋势的形成。

(6) 信息熵值在增大过程中, 是价格变化由趋势变为无序的过程, 一般不是入场的时机<sup>[10]</sup>。信息熵值在  $H_a=1.19$  以下时的变化过程除外。

由以上的几点分析, 就对文中起初所提及的几个问题得到了很好的诠释, 而入场时间、持仓时间、出场时间方面也得到了较好的回答, 在熵值减小过程中,  $H=H_a=1.19$  是比较好的买入入场点,

反之,  $H_a=1.19$  则是较好的卖出离场点, 同样对于  $H$  下穿  $H_b=1.58$  时, 为较好的买入入场点,  $H$  上穿  $H_b=1.58$  时, 卖出离场。

同样, 信息熵值能够多次有效给出价格变化局部行情的转折点, 牛市和熊市对应不同特点的熵值, 熵值增大时往往伴随着价格的震荡或者下跌, 信息熵值的这些规律足以反映复杂市场的一个层面。

参照图 2(A-W) 中对应入场、离场点以及买入、卖出量化收益进行总结, 表 3 即为做空统计表。

表 3 做空出场、入场对应点值及其收益表

| 入场价格/元      | 出场价格/元      | 收益/元    |
|-------------|-------------|---------|
| 1168.43 (A) | 793.25 (B)  | 375.18  |
| 995.72 (C)  | 522.98 (D)  | 472.74  |
| 738.03 (E)  | 553.44 (F)  | 184.59  |
| 1648.47 (I) | 1422.26 (J) | 226.21  |
| 1687.28 (M) | 1650.67 (N) | 36.61   |
| 1736.48 (O) | 1046.16 (P) | 690.32  |
| 4146.30 (S) | 1974.01 (T) | 2172.29 |
| 3046.09 (W) | 2375.77*    | 670.32  |

注: 相应字母为图 2 中对应的时间, 下同; \* 表示的时间为 2012 年 3 月 22 日。

从以上数据可知, 根据指标上穿  $H_b=1.58$  做空入场, 下穿平仓离场, 每一笔交易均盈利, 最后一次是在 2010 年 03 月 18 日做空入场, 截至到 2012 年 03 月 22 日指数为 2375.77, 盈利为 670.32 元。

表 4 做多出场、入场对应点值及其收益表

| 入场价格/元      | 出场价格/元      | 收益/元    |
|-------------|-------------|---------|
| 924.24 (G)  | 1017.13 (H) | 92.89   |
| 1916.25 (K) | 1978.43 (L) | 62.18   |
| 1297.19 (Q) | 5409.40 (R) | 4112.21 |
| 2198.11 (U) | 3471.44 (V) | 1273.33 |

表4为做多统计表. 同理, 指标下穿  $H_a = 1.19$  做多入场, 上穿平仓离场, 每一笔交易也均盈利.

#### 4 结论

借鉴信息熵的概念, 将上证指数一系列数据作为信息, 利用数据每日变化作为信息熵计算的基础, 并利用小波降噪技术实现熵值去噪. 当然, 文中构建的信息熵技术指标阈值适合于长线交易者; 如果要做短线交易, 有4种办法解决时间跨度问题, 一是缩小周期  $Circl$  的值, 二是将计算得到的信息熵值降噪处理幅度减小, 三是可以将前2种方法相结合, 四是可以选取较小周期数据(如4H、1H、30M、15M、5M、1M等), 方法则完全相同. 当然, 不管利用那种方法, 其信息熵阈值的选取一定要有灵活变化.

#### 参考文献:

[1] Cover T M, Thomas J A. 信息论基础[M]. 阮吉寿, 张

华, 译. 北京: 机械工业出版社, 2009.

- [2] 贾世楼. 信息论理论基础[M]. 哈尔滨: 哈尔滨工业大学出版社, 1986.
- [3] 王君彧. 指数现套利建模及交易实现研究[D]. 上海: 复旦大学, 2008.
- [4] 许宏伟, 刘国衡. 交易系统设计方法的探讨[J]. 郑州航空工业管理学院学报, 2001, 19(4): 24-29.
- [5] 侯木舟, 袁修贵. 基于 Matlab 的小波分析在股市技术分析中的应用[J]. 系统工程, 2001, 14(5): 87-92.
- [6] 董长虹. Matlab 小波分析工具箱原理与应用[M]. 北京: 国防工业出版社, 2004.
- [7] 王涛. 投资开放式基金的风险控制研究[D]. 贵阳: 贵州大学, 2008.
- [8] 银河期货. 期货程序化交易系统[EB/OL]. [2011-10-29]. <http://www.yhqh.net/html/72-18/18169.htm>.
- [9] 黎亮. 个人投资者自动交易系统的应用研究[D]. 南宁: 广西大学, 2009.
- [10] 魏凌艳. 我国大豆与豆粕期货跨品种套利交易模型研究[D]. 成都: 西南交通大学, 2008.

## Construction and Application of Information Entropy Index for Securities Market

WANG Bao-shan, SHAO Ming-xing, DING Ai-liang, XU Bing-zhen\*  
( Faculty of Science, Ningbo University, Ningbo 315211, China )

**Abstract:** This article selects the data from each trading day of Shanghai Securities to conduct statistical analysis. First, we use the wavelet technique with noise reduction to obtain the information entropy of technical indicators, and then choose the appropriate information entropy threshold for carrying out the analysis. In so doing, we come up with the corresponding information entropy index of the price movements and market characteristics.

**Key words:** information entropy; wavelet analysis; entropy threshold

(责任编辑 章践立)