

# 证据理论在证券投资组合中的应用

易昆南, 晏玉梅, 徐艳卫

(中南大学 数学科学与计算技术学院, 湖南 长沙 410075)

**摘要:** 针对上班族的一些特点以及我国股票市场的特性, 利用证据理论进行组合投资的分析, 并以实例给出组合投资的比例, 说明证据理论在证券投资组合中有着广泛的应用。

**关键词:** 证据理论; 证券投资; 组合投资

中图分类号: F830.91

文献标志码: A

文章编号: 1673-9833(2009)04-0085-03

## The Application of Evidence Theory in Securities Portfolio

Yi Kunnan, Yan Yumei, Xu Yanwei

(Department of Mathematics, Central South University, Changsha 410075, China)

**Abstract:** Aiming at characteristics of China white-collar workers and China stock markets, analyses portfolio investment with theory evidence. And gives an example of proportions of portfolio investment to expound that evidence theory is widely used in securities portfolio.

**Keywords:** evidence theory; securities investment; portfolio investment

证券组合管理理论最早由美国著名经济学家哈里·马柯威茨于1952年系统提出。在此之前, 偶尔有人也曾在论文中提出过组合的概念, 但经济学家和投资管理者一般仅致力于对个别投资对象的研究和管理。自此之后, 经济学家们一直在利用数量化方法不断丰富和完善组合管理的理论和实际投资管理方法, 并使之成为投资学中的主流理论之一<sup>[1]</sup>。本文针对上班族的一些特点以及我国股票市场的特性, 利用证据理论进行组合投资的分析, 并提出购买建议。

### 1 证据理论概述

证据理论是目前在决策融合研究中受到广泛重视的一种数学方法, 是由Dempster于1967年研究统计问题时首先提出来的, 并给出了上、下概率的概念及其合成规则, 第一次明确给出了不满足可加性的概率, Shafer把它推广到一般的情况, 因此又称为Dempster-

Shafer证据理论(简称D-S推理或证据理论)。文献[2]综合了概率理论的主客观解释, 提出了概率的构造性解释观点, 指出概率是人们在证据的基础上构造出的对一命题为真的信任程度(简称信度)。在证据理论中, 证据不是指实证据, 而是指人们的经验和知识的一部分, 是人们对该问题所作的观察和研究的结果。

证据理论的基本概念及组合规则<sup>[3]</sup>如下。

1) 辨识框架(Frame of discernment)

D-S证据理论中, 一个非常重要的概念就是辨识框架, 它被描述为:  $\div$ 任意证据均可用一个非空集合 $\Theta$ 来表示, 即该证据所支持的命题可用 $\Theta$ 的子集表示, 这个集合被称之为辨识框架,  $2^\Theta$ 称为 $\Theta$ 的幂集。

2) 基本可信度分配(Basic Probability Assignment)

证据对 $\Theta$ 的幂集中元素的支持程度由如下映射关系表示:  $m: 2^\Theta \rightarrow [0, 1]$ ,  $m$ 满足:

$$\sum_{A \in 2^\Theta} \{m(A) | A \subseteq \Theta\} = 1, \text{ 且 } m(\emptyset) = 0;$$

收稿日期: 2009-07-03

作者简介: 易昆南(1954-), 男, 湖南长沙人, 中南大学教授, 硕士生导师, 主要从事随机数学与建模方面的教学与研究,

ykn\_88@sina.com

对任意  $A \in 2^\Theta$ ,  $m(A)$  表示证据支持命题  $A$  发生的程度, 但不表示对任何  $A$  的真子集的支持程度。函数  $m$  称为给定证据下的基本可信度分配函数。

### 3) Dempster 合成公式

证据理论处理不确定性的重要工具是 Dempster 合成公式, 它反映了证据的联合作用。

设  $m_1$  和  $m_2$  是同一辨识框的 2 个证据所对应的基本可信度分配, 焦点分别为  $A_1, A_2, \dots, A_n$  和  $B_1, B_2, \dots, B_n$ , 则由下式定义的函数  $m: 2^\Theta \rightarrow [0, 1]$  是 2 证据联合后的基本可信度分配:

$$m(A) = \begin{cases} 0, & A = \emptyset; \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}, & A \neq \emptyset. \end{cases}$$

上式被称为 Dempster 合成公式, 当多个证据合成时, 可以使用该式级联运算。

## 2 基于证据理论的决策程序

### 1) 建立识别框架

根据专家的经验、知识及其对该问题的观察、研究结果, 作为系统的各种生存状态的证据  $A$ , 把生存状态集合作为识别框架  $\Theta$ 。

### 2) 确定基本可信度分配

根据一批证据  $A$ , 证据  $A$  对  $\Theta$  的幂集中元素的支持程度由如下映射关系表示。

$m: 2^\Theta \rightarrow [0, 1]$ ,  $m$  满足:

$$\sum_{A \in 2^\Theta} \{m(A) | A \subseteq \Theta\} = 1, \text{ 且 } m(\emptyset) = 0;$$

则  $m$  即为识别框架  $\Theta$  上的基本可信度分配。如果有多批证据同时作用于识别框架上, 可通过 Dempster 合成法则求出基本可信度分配  $m$ 。

### 3) 求出各种状态出现的主观概率

由于单点集合上似真度等于众信度, 对于任意  $a \in \Theta$ , 利用定义  $Pl(\{a\}) = \sum_{\{a\} \subset A} m(A)$ , 求出单点似真度  $Pl(\{a\})$ , 然后进行归一化处理, 便于得出主观概率  $P(a)$ 。

### 4) 计算各决策期望值并择优

假设决策集为  $A = \{a_1, a_2, \dots, a_n\}$ , 评估函数为  $f(a_i, x_j)$  ( $i=1, 2, \dots, h; j=1, 2, \dots, k$ ), 则目标决策期望值  $E(a_i) = \sum_{j=1}^k f(a_i, x_j) p(x_j)$ ,  $i=1, 2, \dots, h$ ; 最优目标决策期望值为:  $E(a) = \max\{E(a_1), E(a_2), \dots, E(a_h)\}$ 。

## 3 实例分析

由于上班族不能时时盯着股票, 不适合做短线; 另外, 根据股票的周期性、我国目前股市的频繁震荡特性以及 macd 指标, 最长周期取 26 个交易日, 所以在时间上选择周期为 1 个月, 在股票的选择上根据组合投资分散原则, 选择 6 支股票, 分别来自不同行业, 由于精力、资金量等原因, 从中选择 3 支股票进行组合投资。这 6 支股票分别是: 山东黄金-600547 ( $a_1$ )、中铁二局-600528 ( $a_2$ )、赣粤高速-600269 ( $a_3$ )、八一钢铁-600581 ( $a_4$ )、中海油服-601808 ( $a_5$ )、成发科技-600391 ( $a_6$ ) 和上证指数-000001, 2008 年 3 月 ~ 2009 年 2 月各支股票的收益率见表 1。

表 1 2008 年 3 月 ~ 2009 年 2 月各支股票的收益率  
Table 1 The stock yields from Mar. 2008 to Feb. 2009

时 间	股 票 编 号								
	600547	600528	600269	600581	601808	600391	000001		
2008 年	3 月	-0.339 3	-0.190 6	-0.150 6	-0.186 8	-0.328 6	-0.355 5	-0.201 4	
	4 月	-0.158 4	-0.071 3	0.029 0	0.299 9	0.187 4	0.121 3	0.063 5	
	5 月	0.023 2	-0.341 5	-0.165 7	-0.130 3	-0.048 2	-0.004 7	-0.070 7	
	6 月	-0.152 6	-0.239 6	-0.213 0	-0.193 1	-0.050 7	-0.368 0	-0.203 1	
	7 月	0.042 9	0.029 5	0.061 3	-0.012 5	-0.064 5	0.234 1	0.014 5	
	8 月	-0.334 1	-0.299 2	-0.129 9	-0.406 2	-0.269 4	-0.276 1	-0.136 3	
	9 月	0.053 3	-0.011 7	0.100 7	0.057 6	-0.033 1	0.001 4	-0.043 2	
	10 月	-0.148 8	0.106 5	-0.201 0	-0.152 4	-0.391 1	-0.424 9	-0.246 3	
	11 月	0.213 9	0.474 2	0.002 5	0.344 0	0.169 9	0.186 8	0.082 4	
	12 月	0.243 9	0.009 7	-0.021 3	-0.049 6	0.078 9	0.256 3	-0.026 9	
	2009 年	1 月	-0.053 7	0.117 2	0.089 7	0.187 3	0.157 3	0.220 9	0.093 3
		2 月	0.514 0	-0.051 4	0.025 9	0.054 9	-0.056 7	0.148 7	0.046 3

令  $A_{ij}$  表示第  $j$  只股票第  $i$  月的收益率 ( $i=1, 2, \dots, 12; j=1, 2, \dots, 6$ ), 则第  $j$  只股票的期望收益为

$$E(A_{ij}) = \sum_{i=1}^{12} A_{ij} / 12; \text{ 组合收益为 } E(B) = \sum_{i=1}^6 w_i E(A_{ij})^{[4]}, w_i$$

是各个股所占比例。证据理论强调依据证据为一个命题赋予真值, 把所能认识到的所有可能结果集合作为识别框架。在证券组合投资中, 可把所选择的证券作为识别框架, 本文把集合  $\{a_1, a_2, a_3, a_4, a_5, a_6\}$  作为识别框架, 由每只股票所获得的信息量作为识别框架的证据  $A$ 。从 6 只股票中任取 3 只进行组合投资, 可把集合  $\{a_k, a_m, a_n\}$  作为决策集。

由于证券本身是风险与收益的组合<sup>[5]</sup>, 即包括高风险高收益、高风险中等收益、高风险低收益、低风险低收益、低风险中等收益、低风险高收益(存在于某些债券)等几种类型, 所以证券组合投资问题是多目标决策问题。虽然不能确定 3 只股票的投资比例, 但有一定把握确定其中某一部分股票集合所占的比例。因此, 可根据聚类分析的相关原理, 按照不同属性, 把所要研究对象分成不同类集合。比如可把投资者分为冒险型、稳健型、保守型 3 类, 对于保守型投资者来说, 他投资于低风险低收益的股票比例要相对大一些, 而投资于高风险高收益的股票比例要相对小一些。但是对于冒险型投资者来说, 情况则相反, 他会把很大比例资金投资于风险大且收益高的股票上。而稳健型投资者的投资行为介于以上两者之间。现在假设某个投资者把资金分别按 1/3、2/3 的比例投资于风险中等偏低和风险中等偏高的 2 类股票上, 那么根据证据理论, 可以建立 mass 函数。在这些股票中, 根据证据  $E$ , 在框架上产生的基本可信度分配  $m$  为:

$$m(A) = \begin{cases} 1/3, & A = \{a_1, a_2, a_3, a_4\}, \\ 2/3, & A = \{a_3, a_4, a_5, a_6\}, \\ 0, & A = \text{其它}, \end{cases}$$

则有:  $Pl(\{a_1\}) = 1/3$ ,  $Pl(\{a_2\}) = 1/3$ ,  $Pl(\{a_3\}) = 1$ ,  $Pl(\{a_4\}) = 1$ ,  $Pl(\{a_5\}) = 2/3$ ,  $Pl(\{a_6\}) = 2/3$ 。所以

$$P(a_1) = Pl(\{a_1\}) / \sum_{i=1}^{12} Pl(\{a_i\}) = \frac{1}{12}, \text{同理可得:}$$

$$P(a_2) = \frac{1}{12}, P(a_3) = \frac{1}{4}, P(a_4) = \frac{1}{4}, P(a_5) = \frac{1}{6}, P(a_6) = \frac{1}{12}。$$

若选择  $a_1$ 、 $a_2$ 、 $a_3$  3 支股票进行组合投资, 还需确定投资于这 3 支股票的比例系数, 此时只需对  $P(a_1)$ 、 $P(a_2) = 1/12$ 、 $P(a_3)$  进行归一化处理, 即可得到  $P(a_1) = 1/5$ ,  $P(a_2) = 1/5$ ,  $P(a_3) = 3/5$ , 则  $a_1$ 、 $a_2$ 、 $a_3$  组合投资的期望收益为:  $E(a_1a_2a_3) = -0.0380$ 。

同理可得决策集合数值, 整个决策集合数值如下:

$$E(a_1a_2a_3) = -0.0380, E(a_1a_2a_4) = -0.0188,$$

$$E(a_1a_2a_5) = -0.0388, E(a_1a_2a_6) = -0.0226,$$

$$E(a_1a_3a_4) = -0.0283, E(a_1a_3a_5) = -0.0432,$$

$$E(a_1a_3a_6) = -0.0324, E(a_1a_4a_5) = -0.0272,$$

$$E(a_1a_4a_6) = -0.0163, E(a_1a_5a_6) = -0.0376,$$

$$E(a_2a_3a_4) = -0.0323, E(a_2a_3a_5) = -0.0215,$$

$$E(a_2a_3a_6) = -0.0408, E(a_2a_4a_5) = -0.0323,$$

$$E(a_2a_4a_6) = -0.0215, E(a_2a_5a_6) = -0.0381,$$

$$E(a_3a_4a_5) = -0.0373, E(a_3a_4a_6) = -0.0291,$$

$$E(a_3a_5a_6) = -0.0421, E(a_4a_5a_6) = -0.0283,$$

那么,  $\max(\{a_k a_m a_n\})$  的决策期望值为  $a_1 a_4 a_6$  的决策期望值。也就是说, 应选取 600547、600581、600391 这 3 只股票进行组合投资, 且投资比例为 1 : 3 : 2。

## 4 结语

从上面的数据可看出, 购买 600547、600581、600391 这 3 支股票, 且采用投资比例为 1 : 3 : 2 时收益最好, 由于采用的数据都是以前的, 在操作中明显滞后, 如何使其具有利用价值, 这就需要对更多的股票进行统计, 然后进行聚类分析, 因为历史会重演<sup>[1]</sup>, 所以利用前面的数据就可以进行有效的应用。

## 参考文献:

- [1] 中国证券协会. 证券投资分析[M]. 北京: 中国财政经济出版社, 2008: 279-281, 220-221.  
Securities Association of China. An Analysis of Securities Investment[M]. Beijing: China Financial & Economic Publishing House, 2008: 279-281.
- [2] Shafer G. A Mathematical of Evidence[M]. Princeton: Princeton University Press, 1976: 9-18.
- [3] 倪明, 单渊达. 证据理论及其应用[J]. 电力系统自动化, 1996(3): 76-80.  
Ni Ming, Shan Yuanda. Evidential Theory and Its Applications[J]. Automation of Electric Power Systems, 1996(3): 76-80.
- [4] 中国证券协会. 证券投资基金[M]. 北京: 中国财政经济出版社, 2008: 237-240.  
Securities Association of China. Securities Investment Funds [M]. Beijing: China Financial & Economic Publishing House, 2008: 237-240.
- [5] 戴维·G·卢恩伯格. 投资科学[M]. 北京: 中国人民大学出版社, 2005: 308-309.  
David G Luenberger. Investment Science[M]. Beijing: China Renmin University Press, 2005: 308-309.

(责任编辑: 廖友媛)