

文章编号: 1000-5013(2009)02-0221-04

# 马尔可夫链在前沿资产组合中的运用

邱世斌, 陈燕武

(华侨大学 商学院, 福建 泉州 362021)

摘要: 将马尔可夫链与资产前沿组合有效结合,以福建板块的5支股票数据做为样本,运用马尔可夫链和自相关性来预测股票在将来的收益率及其分布情况.根据马可维茨理论得出股票市场的前沿组合,对中国股票市场的实例进行实证分析.在择优标准的选取过程中,单位收益的风险最小或是单位风险的收益最大只是一种标准,其选择的方法可以依据相关理论进行选择.

关键词: 马尔可夫链; 自相关性; 前沿资产组合; 择优标准

中图分类号: O 211.62; F 224.0

文献标识码: A

按照证券市场的运行机制,将交易时间离散化为  $t = 1, 2, 3, \dots$ . 在不同的交易时间点,证券所处的状态可以用随机变量来表示,且每一种不同的状态都有相应的概率,称为状态概率.当证券从一种状态转移到另一种状态时,存在着转移的概率,称为转移概率.如果转移概率只与目前相邻的状态变化有关,那么这种离散状态按照离散时间的随机转移系统过程,称为马尔可夫过程<sup>[1]</sup>.在一些系统中,将来的状态的发生不仅与前一期的状态有关,而且与前几期的状态都有关.这就需要对马尔可夫链的基本原理进行拓展.本文将马尔可夫链与资产前沿组合有效结合,对中国股票市场的实例进行实证分析.

## 1 股票的收益率及分布情况

随机从股票市场的福建板块抽取出5支股票,作为样本来预测在下一时期各自出现状态的概率.这5支股票分别为中国武夷、福建高速、福建水泥、金龙汽车、七匹狼.样本数据期为2005年11月22日至2007年3月18日,目的是要预测2007年3月19日各股票出现不同状态的概率.数据来源于兴业证券行情系统,并经过一些必要的处理.

首先,以收益率  $r$  为划分状态 ( $S_i$ ) 的指标,依据其大小来划分其所属的状态.5支股票的状态<sup>[2]</sup> 收益率  $r$  分别:  $r \leq -0.0212$ ,  $-0.0212 < r \leq -0.0039$ ,  $-0.0039 < r \leq 0.0104$ ,  $0.0104 < r \leq 0.0236$ ,  $r > 0.0236$ . 然后,依据收益率所处的状态,将各个样本股票依时间顺序排列,以中国武夷这支股票为例,其在样本数据期的状态序列(该序列是以状态的下标表示)为

4322233212524333442313533242543534324343224144414432133425433313212235342352422443  
22335354351532552115415533114355115255154521415341414224531425245352214523313323222113  
43324112132152244235423133353543332451355343

用  $N\{\sum_{i=1}^{T-1} X_{it} = S_i\}$  表示在  $t$  期以前,出现状态  $S_i$  的次数;用  $N\{\sum_{i=2}^{T-1} X_{it} = S_i/X_{t-1} = S_j, i, j = 1, 2, 3, 4, 5\}$  表示在  $t$  期以前,从状态  $S_j$  转移到状态  $S_i$  的次数;用  $N\{\sum_{i=3}^{T-1} X_{it} = S_i/X_{t-1} = S_j, X_{t-2} = S_k, i, j, k = 1, 2, 3, 4, 5\}$  表示在  $t$  期以前,从状态  $S_k$  转移到状态  $S_j$ , 然后再转移到状态的次数;用  $F\{\sum_{i=1}^{T-1} X_{it} = S_i\}$  表示在  $t$

收稿日期: 2008-01-18

通信作者: 陈燕武(1971-),女,副教授,主要从事金融工程和金融计量的研究. E-mail: lzscyw@hqu.edu.cn.

基金项目: 教育部高等学校博士点学科专项科研基金资助项目(20050385001);福建省高等学校新世纪优秀人才支持计划项目(07FJRC07);华侨大学科研基金资助项目(06BS211)

期以前, 出现状态  $S_i$  的频率; 用  $F\{\sum_{i=2}^{T-1} X_t = S_i / X_{t-1} = S_j, i, j = 1, 2, 3, 4, 5\}$  表示在  $t$  期以前, 从状态  $S_j$  转移到状态  $S_i$  的频率; 用  $F\{\sum_{i=3}^{T-1} X_t = S_i / X_{t-1} = S_j, X_{t-2} = S_k, i, j, k = 1, 2, 3, 4, 5\}$  表示在  $t$  期以前, 从状态  $S_k$  转移到状态  $S_j$ , 然后再转移到状态  $S_i$  的频率.

通过自相关分析,  $t$  期出现的状态与  $t-1$  期和  $t-2$  期存在着显著的自相关性. 从  $t-3$  期后, 其自相关性不显著. 说明  $t$  期出现各种状态的概率不仅与  $t-1$  期有关, 而且还与  $t-2$  期有关, 即将来出现各种状态的概率依据于过去各时期、各状态所发生的概率. 文中所用样本属大样本. 在大样本情况下, 样本的频率近似等于总体的概率. 以  $P\{X_t = S_i\}^*$  来表示过去各时期状态发生的概率, 有

$$P\{X_t = S_i\}^* = F\{\sum_{i=1}^{T-1} X_t = S_i\} = \frac{N\{\sum_{i=1}^{T-1} X_t = S_i\}}{\sum_{i=1}^5 N\{\sum_{i=1}^{T-1} X_t = S_i\}}, \quad (1)$$

其结果如表 1 所示. 表 1 中的小数均取小数点后 3 位, 但计算时按整个数来计算,  $X_t$  为状态,  $N$  为频数,  $F$  为频率.

表 1 概率统计结果

Tab. 1 Statistical results of the probabilities

$X_t$	1	2	3	4	5
$N$	31	46	58	41	36
$F$	0.146	0.217	0.274	0.193	0.170

运用马尔可夫链的单项转移, 可得  $P\{X_i = S_i / X_{t-1} = S_j\}$ , 即第  $t$  期出现的

不同状态的概率与前期出现的状态有关. 由于预测期的前一期状态为  $S_3$ , 因此, 只需要计算  $P\{X_i = S_i / X_{t-1} = S_3\}$  即可. 该样本属于大样本, 其转移的概率近似等于第  $t$  期以前转移的频率. 即

$$P\{X_t = S_i / X_{t-1} = S_3\} = F\{\sum_{i=2}^{T-1} X_t = S_i / X_{t-1} = S_3, i = 1, 2, 3, 4, 5\} = \frac{N\{\sum_{i=2}^{T-1} X_t = S_i / X_{t-1} = S_3, i = 1, 2, 3, 4, 5\}}{\sum_{i=1}^5 N\{\sum_{i=2}^{T-1} X_t = S_i / X_{t-1} = S_3, i = 1, 2, 3, 4, 5\}}. \quad (2)$$

通过计算, 其单项转移概率( $P_1$ )如表 2 所示. 表 2 中, “ $i/j$ ”表示从状态  $i$  转到状态  $j$ .

表 2 马尔可夫链的转移概率

Tab. 2 Transforming probabilities of Markov chain

$P_1$			$P_2$			$S_i$	$P_i$
$X_t$	$N$	$F$	$X_t$	$N$	$F$		
1/3	6	0.105	1/4, 3	0	0	1	0.108
2/3	14	0.246	2/4, 3	5	0.385	2	0.255
3/3	16	0.281	3/4, 3	4	0.308	3	0.282
4/3	8	0.140	4/4, 3	1	0.077	4	0.156
5/3	13	0.228	5/4, 3	3	0.231	5	0.199

同理, 可以得到相类似的马尔可夫链的二项转移. 即表示从  $t$  期的前两期的状态转移到将来各个状态的概率, 其二项转移概率( $P_2$ )如表 2 所示. 在表 2 中, “ $i/j, k$ ”表示从状态  $k$  转到状态  $j$  再转到状态  $i$ . 因为  $t$  期出现不同状态的概率与前两期存在着自相关性, 在该前提下, 对各项转移分别赋予 1/2, 1/3, 1/6 的权重, 可以得到  $t$  期各状态的概率为

$$P\{X_t = S_i\} = 1/2 \times P\{X_t = S_i\}^* + 1/3 P\{X_t = S_i / X_{t-1} = S_3\} + P\{X_t = S_i / X_{t-1} = S_3, X_{t-2} = S_4\}. \quad (3)$$

各状态的最终转移概率( $P_i$ ), 如表 2 所示. 同理, 可以得到其他股票在  $t$  期各状态的概率分布, 结果如表 3 所示<sup>[3]</sup>.

表 3 样本股的概率分布

Tab. 3 Probability distribution of sample stocks

$S_i$	七匹狼	福建高速	金龙汽车	福建水泥
1	0.223	0.172	0.116	0.218
2	0.205	0.247	0.217	0.133
3	0.236	0.268	0.296	0.263
4	0.119	0.195	0.191	0.231
5	0.217	0.117	0.180	0.156

在股票市场中, 投资者所关心的是各支股票的收益率( $r$ ), 而收益率在状态区间内的变动是在可以接收的范围. 因为在状态区间内收益率的变化都较小, 而在状态区间发生了变动, 投资者就会改变其投资组合. 把属于某一状态的收益率加以简单的加权平均, 就可以得到样本期间各状态的平均收益率  $\bar{r}$ . 经计算, 每支股票的平均收益率  $\bar{r}$ 、期望收益率  $E(X)$  和方差  $D(X)$ , 如表 4 所示.

## 2 股票的前沿组合

表 4 样本股的收益率和方差

Tab. 4 Return rates and variance of sample stocks

参数	中国武夷	七匹狼	福建高速	金龙汽车	福建水泥	
$r$	1	- 0. 045	- 0. 036	- 0. 046	- 0. 047	- 0. 048
	2	- 0. 011	- 0. 011	- 0. 011	- 0. 012	- 0. 012
	3	0. 003	0. 003	0. 003	0. 003	0. 003
	4	0. 016	0. 016	0. 017	0. 017	0. 016
	5	0. 045	0. 044	0. 037	0. 042	0. 042
$E(X)$	0. 004 5	0. 001 9	- 0. 002 1	0. 003 7	- 0. 000 8	
$D(X)$	0. 000 7	0. 000 8	0. 000 6	0. 000 6	0. 000 9	

通过有效的投资组合, 可以分散市场中的非系统性风险, 但无法分散系统性风险. 在市场中, 投资者希望在收益一定时, 通过股票组合把自己所承担的风险降到最低, 或是希望在风险一定的情况下, 收益达到最大. 投资者对收益的偏好, 常表现为对预期平均收益的偏好, 对风险的厌恶表现为对方差的厌恶. 因为方差表现为对平均收益率的偏离, 即未来有产生不好结果的可能性<sup>[4]</sup>.

股票投资者通过投资组合, 希望达到预期收益一定、方差最小和收益最大的模型<sup>[5]</sup> 分别为

$$\min w^T V w = \alpha_p, \quad w^T e = E(r_p), \quad w^T I = 1. \quad (4)$$

$$\min w^T e = E(r_p), \quad w^T V w = \alpha_p, \quad w^T I = 1. \quad (5)$$

式(4), (5)中,  $e$  为股票的期望收益率向量,  $E(r_p)$ ,  $\alpha_p$  分别为股票组合  $p$  的期望收益率和标准差,  $w$  为投资于这些股票的投资比例,  $I$  为  $N$  (组合股票的数目) 维单位列向量,  $V$  为各股票间的协方差矩阵.

依据 5 支股票各状态的收益率及概率分布, 其协方差矩阵如表 5 所示. 协方差矩阵  $cov(x, y) = \alpha_x \alpha_y R_{x,y}$ . 其中,  $R_{x,y}$  为  $x, y$  股票收益率的相关系数. 运用 SPSS 软件做相关性分析, 得到各股票收益率间的相关系数, 如表 5 所示. 从表 5 可知, 各股票收益率间的相关系数都是显著的.

表 5 各股票收益率间的协方差矩阵和相关系数

Tab. 5 Covariance matrix and correlation coefficients of the return rates among different stocks

参数	股票	中国武夷	七匹狼	福建高速	金龙汽车	福建水泥
$cov(x, y)$	中国武夷	0. 000 68	0. 000 26	0. 000 13	0. 000 07	0. 000 21
	七匹狼	0. 000 26	0. 000 77	0. 000 05	0. 000 18	0. 000 20
	福建高速	0. 000 13	0. 000 49	0. 000 61	0. 000 07	0. 000 11
	金龙汽车	0. 000 65	0. 000 18	0. 000 07	0. 000 64	0. 000 13
	福建水泥	0. 000 21	0. 000 20	0. 000 11	0. 000 13	0. 000 85
$R_{x,y}$	中国武夷	1. 000	0. 360	0. 193	0. 099	0. 277
	七匹狼	0. 360	1. 000	0. 071	0. 249	0. 245
	福建高速	0. 193	0. 071	1. 000	0. 116	0. 158
	金龙汽车	0. 099	0. 249	0. 116	1. 000	0. 182
	福建水泥	0. 277	0. 245	0. 158	0. 182	1. 000

将以上的相关数据带入到两个目标方程中去, 并利用拉格朗日定理求解, 所得出的投资组合就是股票的前沿投资组合. 计算后可以得到一个关于  $\sigma^2(r_p)$  与  $E(r_p)$  的方程<sup>[6]</sup>, 即

$$\frac{\sigma^2(r_p)}{0. 000 242} - \frac{(E(r_p) - 0. 001 333)^2}{0. 000 016} = 1, \quad (6)$$

$$w_p = \left[ \begin{matrix} 0. 038 & 0. 172 & 0. 426 & 0. 160 & 0. 203 \end{matrix} \right]^T + \left[ \begin{matrix} 110. 333 & - 22. 206 & - 103. 767 & 74. 162 & - 58. 522 \end{matrix} \right]^T E(r_p). \quad (7)$$

只需给定一个期望的收益率, 就可以通过一定的股票投资组合达到方差最小, 并计算出相应的投资比例. 当然投资比例也可能出现负数, 则表示的股票是卖空股票. 虽然, 在中国证券市场上不能进行卖空, 但目前中国股票市场中存在认沽权证和认购权证, 其功能可以抵补不能卖空的缺点. 投资者要选择什么样的收益率, 应该看投资者对待风险的态度. 如果投资者能够承受高风险, 那可以选择较高的投资组合收益率; 反之同理. 将上式方程表示为一条双曲线, 如图 1 所示. 图 1 中,  $\alpha_1$  为投资的最小风险, 但是它所代表的收益也是最小的( $E_1$ ). 当投资者选择高收益率时, 也承担了更高的风险水平. 图 1 中,  $E_2 > E_1$ , 但组合 2 的风险也大于组合 1 的风险, 即  $\alpha_2 > \alpha_1$ . 当然, 投资者不会选择  $A$  点以下的投资组合. 这意味着在相同的风险水平下, 收益率比  $A$  点以上的前沿组合的收益率更小. 在相同的风险水平下, 一个理性的投资者只会选择有较高收益率的投资组合.

### 3 投资的组合的择优标准

图 1 中, 上半部分是有效的前沿资产组合, 投资者只会在上半部分选择投资组合. 因为在同样的风险水平下, 上半部分表示的收益大于下半部分的收益. 理论上, 最好的组合点应该是投资者的效用函数与投资函数相切的点. 设投资者的效用函数为  $u(E(r_p), \sigma_p)$ , 投资者的投资点为图 1 中的 B 点. 但是, 每个投资者的效用函数都是不同的, 无法进行一般性的分析, 故很难找到一种普遍适用的效用函数. 上面模型所描述的是在收益一定的情况下风险最小. 从相同的角度出发, 让单位风险收益最大, 或是单位收益的风险最小的择优标准, 即可得到一个比较好的投资组合.

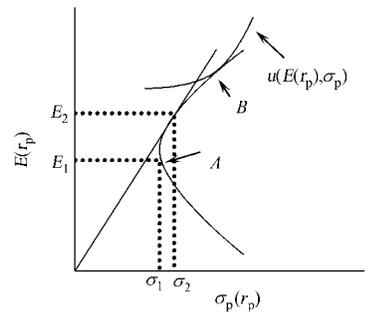


图 1 期望收益率和股票组合的标准差之间的关系

Fig. 1 Relationships between the return rate and standard deviation of portfolios

用数学模型来描述该择优组合, 即

$$\min(\alpha(r_p)/E(r_p)), \quad \max(E(r_p)/\sigma(r_p)), \quad (8)$$

$$\frac{\alpha(r_p)}{0.000242} - \frac{(E(r_p) - 0.001333)^2}{0.000016} = 1. \quad (9)$$

用图形表示出来, 就是原点与双曲线上的点. 与双曲线相切的点就是所得到最优的投资组合点, 即图 1 中的  $(\sigma_2, E_2)$  这一点.

### 4 结束语

尽管模型是以 5 支福建板块的股票数据为样本来研究的, 但具有普遍的适用性, 可以用于整个股票市场中的股票投资组合. 虽然只讨论股票市场, 该模型也可推广至债券市场, 特别是加入了国债这种无风险资产后, 只要在组合中再引入市场组合, 就可以得到资本资产定价模型. 在该模型中都不考虑股票的交易费用, 但是在现实中交易费用不得不考虑在内. 因此, 可以从资产组合的期望收益率中扣除费用率, 就可以把该变量引入到模型.

#### 参考文献:

- [1] 唐向来. 马尔可夫链在个股分析中的应用[J]. 湖南科技学院学报, 2005, 26(11): 26-29.
- [2] 关丽娟, 赵鸣. 沪综指走势的马尔可夫链模型预测[J]. 山东行政学院学报, 2005(8): 101-102.
- [3] 陈祥钟, 黄荣坦. 中国股市波动率变化特征的实证分析[J]. 华侨大学学报: 自然科学版, 2006, 27(4): 437-440.
- [4] 范良铸. 中国证券市场的最优投资组合选择研究[D]. 武汉: 武汉大学, 2005.
- [5] HUANG Chi-fu. 金融经济学基础[M]. 宋逢明, 译. 北京: 清华大学出版社, 2003.
- [6] STAMPFLI J, GOODMAN V. 金融数学[M]. 蔡明超, 译. 北京: 机械工业出版社, 2004.

## Application of Markov Chain in the Frontier Portfolio

QIU Shi-bin, CHEN Yan-wu

(College of Commerce, Huaqiao University, Quanzhou 362021, China)

**Abstract:** This paper effectively analyses the Chinese stock market with the Markov chain and frontier portfolio by taking the data of five stocks of Fujian Province's sectors as samples; and we forecast the future return rate and distribution of the stocks; then the frontier portfolio of the stock market is concluded from the Markowitz theory. In the process of choosing criteria, the concept of maximum unit risk return and the minimum unit return risk is one of the selection criteria; the ways of choosing criterion can be carried through according to the related theories.

**Keywords:** Markov chain; autocorrelation; frontier portfolio; selection criterion

(责任编辑: 钱筠 英文审校: 司福成)