

基于Markov模型对股票的价格预测

王 洲, 范爱华*

安徽工业大学商学院, 安徽 马鞍山

收稿日期: 2022年7月11日; 录用日期: 2022年7月22日; 发布日期: 2022年8月2日

摘 要

股票是金融市场上重要的金融工具, 正确的股票预测能够帮助投资者做出更好的决策, 同时股票价格波动具有很好的马氏性。本文选取深圳A股“平安银行(SZ000001)”45个交易日的数据, 通过数据整理获取股票价格的概率转移矩阵并建立马尔可夫模型, 运用模型分析并对股价运动进行预测, 得到了较为理想的结果。说明马尔可夫模型对于股票的预测具有可行性, 投资者能够结合实际情况在股票市场做出更好的决策。

关键词

股票市场, 价格预测, 马尔可夫链, 概率转移

Stock Price Prediction Based on Markov Model

Zhou Wang, Aihua Fan*

School of Business, Anhui University of Technology, Ma'anshan Anhui

Received: Jul. 11th, 2022; accepted: Jul. 22nd, 2022; published: Aug. 2nd, 2022

Abstract

The stock is an important financial instrument in the financial market. The correct stock prediction can help investors make better decisions, and the stock price fluctuation has a good Markov property. This paper selects the data of 45 trading days of Shenzhen A-share "Ping An Bank (SZ000001)", obtains the probability transfer matrix of stock price through data collation, and constructs the Markov model, and predicts the stock price movement by model analysis, and obtains ideal results. It shows that Markov model is feasible for stock prediction and investors can make better decisions in the stock market according to the actual situation.

*通讯作者。

Keywords

Stock Market, Price Forecast, Markov Chain, Probability Transfer

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

股票市场是市场经济中最重要的资本市场, 2021年9月3日北京证券交易所成立, 标志着中国的股票市场自20世纪90年代初重新建立以来即将进入一个新的发展阶段, 股票市场是企业直接融资的主要渠道, 走市场化的经济发展道路, 必须充分地规范发挥股票市场作用。股价预测方法有多种, 多年来相关学者提出了各种预测方法, 总结出来有三类: 基于统计学方法的股票价格波动预测模型、机器学习预测模型和基于随机过程的预测模型[1]。考虑到经济环境是不断变化的, Markov模型正是通过对事物不同状态的初始概率与状态之间转移概率的研究, 确定状态变化趋势, 预测事物的未来。目前, 马尔科夫模型主要应用于产品市场占有率、市场销售情况、期望利润等方面的预测, 本文引入运用状态划分方法, 借助C-K方程以及平稳分布, 预测股价未来所处状态以及所处区间, 并与真实数据进行对比, 通过在实际股价预测中的应用, 为投资者在实际决策中提供更为可靠的依据。

2. 理论基础

2.1. Markov链的定义

定义1 [2]: 随机过程 $\{X_n, n=0,1,2,\dots\}$ 称为Markov链, 若它只取有限或可列个值, 对任意 $n \geq 0$ 及任意状态 $i, j, i_0, i_1, \dots, i_{n-1}$, 有:

$$P\{X_{n+1}=j | X_0=i_0, X_1=i_1, X_{n-1}=i_{n-1}, X_n=i\} = P\{X_{n+1}=j | X_n=i\}$$

上式刻画了Markov链的特性, 称为Markov性, 也称无后效性, 其是指事物在以后的状态只与本阶段的状态有关, 而与以前任何阶段所处的状态均无关。

2.2. 转移概率

定义2 [2]: 将 $P_{ij} (i, j \in S)$ 排成一个矩阵形式, 令:

$$P = (P_{ij}) = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \cdots \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

称 P 为转移概率矩阵, 一般简称为转移矩阵。由于概率是非负的, 且过程必须转移到某种状态, 所以容易看出 $P_{ij} (i, j \in S)$ 有性质:

- 1) $P_{ij} \geq 0, i, j \in S$;
- 2) $\sum_{j \in S} P_{ij} = 1, \forall i \in S$ 。

2.3. n 步转移概率和 C-K 方程

定义 3 [2]: n 步转移概率: 称条件概率 $P_{ij}^{(n)} = P\{X_{m+n} = j | X_m = i\}$, $i, j \in S, m \geq 0, n \geq 1$ 为 Markov 链的 n 步转移概率, 相应的称 $P^{(n)} = (P_{ij}^{(n)})$ 为 n 步转移矩阵。

C-K 方程: 对一切 $m, n \geq 0, i, j \in S$ 有:

- 1) $P_{ij}^{(m+n)} = \sum_{k \in S} P_{ik}^{(m)} P_{kj}^{(n)}$
- 2) $P^{(n)} = P \cdot P^{(n-1)} = P \cdot P \cdot P^{(n-2)} = \dots = P^n$

2.4. 平稳分布与极限分布

定义 4 [3]: 设 $\{X_n, n \geq 0\}$ 是齐次马尔可夫链, 状态空间为 I , 转移概率为 P_{ij} , 若:

$$\begin{cases} \pi_j = \sum_{i \in I} \pi_i P_{ij} \\ \sum_{j \in I} \pi_j = 1, \pi_j \geq 0 \end{cases}$$

即若概率分布为 $\pi = \{\pi_1, \pi_2, \dots, \pi_n\}$, 转移概率有 $\pi = \pi P$, 称概率分布 $\{\pi_j, j \in I\}$ 为马尔可夫链的平稳分布。对于平稳分布有 $\pi = \pi P = \pi P^2 = \pi P^n$ 。

称 Markov 链是遍历的, 如果所有状态相通且均是周期为 1 的正常返状态, 对于遍历的 Markov 链, 极限 $\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi_j, j \in I$ 称为 Markov 链的极限分布。

3. 模型分析

3.1. 马氏性分析

股票价格、股票价格区间以及成交量的状态是一族依赖于时间 t 的随机变量, 其变化过程是一个随机过程。且在时间 t 所处的状态只与在时刻 t 的状态有关, 而与时刻 t 以前所处的状态无关, 即具有无后效性。无论从什么时候开始, 股票状态变化过程都保持一种时间历程的不变性, 其一步转移概率只与时间差有关, 而与时间起点无关, 所以转移概率矩阵保持稳定不变。同时股票价格、股票价格区间以及成交量只能产生可列个状态, 且用同一标准划分的各状态应相互独立, 并包含全部可能出现的状况。因此股票价格、股票价格区间以及成交量变化过程符合马尔科夫过程, 可用 Markov 模型来预测股票未来走势。

3.2. 股价分析

现考察一只股票在第 n 天的收盘价格 $X_n (n = 0, 1, 2, 3, \dots)$, 则 $X_n \in [0, +\infty)$, 在 $X_n \in [0, +\infty)$ 中插入 $m-1$ 个分点 $0 < X_1 < X_2 < \dots < X_{m-1}$, 分别记 X_0 和 X_m 为 0 和 $+\infty$, 那么在 $[0, +\infty)$ 被分成 m 个区间, 其中第 i 个区间为 $[X_{i-1}, X_i) (1 \leq i \leq m)$, X_n 只能落在 m 个区间中的某一个里。同时规定: $X_n(\omega) = i$, 当 $\omega \in [X_{i-1}, X_i)$ 时, 我们称股价的状态为 i 。记 P_{ij} 为股票从状态 i 经过一个时间间隔转移到状态 j 的概率, $P_{ij}^{(n)}$ 表示状态 i 经过 n 步转移到状态 j 的概率。由此股票价格的一次转移概率为:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

由 C-K 方程可知 $P(n) = P^n$ 述了股价从一个状态到下一个状态的概率分布状况, 因此, 通过计算 $P(n)$, 我们可以预测 n 个交易日后的股价。即通过比较 $P(n)$ 中第 i 行各项值的大小就可以判断 n 个时间间隔后股价最可能达到的状态。

4. 股价运动马尔可夫链的实证分析

4.1. 以股票收盘价的状态为对象进行预测

以平安银行(SZ000001) 2022年3月01日至5月09日共45个收盘交易日的收盘交易价格资料为例, 运用状态划分的方法, 将各日的收盘价格分为上升、持平、下降三种状态进行分析预测, 统计结果见表1。

Table 1. Statistics of Ping An Bank's closing price

表 1. 平安银行收盘价状态统计

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 状态 | 上升 | 下降 | 上升 | 下降 | 下降 | 下降 | 下降 | 上升 | 上升 |
| 序号 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 状态 | 下降 | 下降 | 上升 | 上升 | 上升 | 下降 | 上升 | 下降 | 上升 |
| 序号 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 状态 | 下降 | 下降 | 下降 | 上升 | 上升 | 上升 | 上升 | 下降 | 上升 |
| 序号 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| 状态 | 下降 | 下降 | 下降 | 上升 | 上升 | 下降 | 下降 | 上升 | 下降 |
| 序号 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 |
| 状态 | 上升 | 下降 | 下降 | 上升 | 持平 | 下降 | 持平 | 下降 | 下降 |

表中可知, 由上升转变为上升状态的次数为7次, 则以概率代替频率, 即由状态 i_1 到状态 i_1 的转移概率: $P_{11} = 7/19 = 0.3684$ 次数为11次, 即 $P_{13} = 11/19 = 0.5790$; 由上升转变为持平状态的次数为1次[4], 即 $P_{12} = 1/19 = 0.0526$; 以此类推, 可以得到 $P_{21} = 0$, $P_{22} = 0$, $P_{23} = 1$, $P_{31} = 11/23 = 0.4783$, $P_{32} = 1/23 = 0.0434$; $P_{33} = 11/23 = 0.4783$, 将其各转移状态情况列成如下表2:

Table 2. Transfer state probability statistics

表 2. 转移状态概率统计

| 状态 | 上升 i_1 | 持平 i_2 | 下降 i_3 |
|----------|----------|----------|----------|
| 上升 i_1 | 0.3684 | 0.0526 | 0.5790 |
| 持平 i_2 | 0 | 0 | 1 |
| 下降 i_3 | 0.4783 | 0.0434 | 0.4783 |

由表2可以得到平安银行的收盘状态转移矩阵:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} = \begin{bmatrix} 0.3684 & 0.0526 & 0.5790 \\ 0 & 0 & 1 \\ 0.4783 & 0.0434 & 0.4783 \end{bmatrix}$$

矩阵 P 中每行即为某一状态的各种转移概率, 故有 $\sum_{j=1}^3 P_{ij} = 1$, $i = 1, 2, 3$ 。

根据表1, 由于所统计数据最后一天5月09日处于下降状态, 而后续无资料, 所以可以认为其初始状态向量 $\pi(0) = (1, 0, 0)$ 利用该向量和状态转移概率矩阵来预测各个收盘日价格所处状态的概率:

第 10 日收盘价状态概率向量:

$$\pi(1) = \pi(0)P = (1, 0, 0) \begin{bmatrix} 0.3684 & 0.0526 & 0.5790 \\ 0 & 0 & 1 \\ 0.4783 & 0.0434 & 0.4783 \end{bmatrix} = (0.3684, 0.0526, 0.5790)$$

第 11 日收盘价状态概率向量: $\pi(2) = \pi(1)P = (0.4127, 0.0445, 0.5428)$ 。

第 12 日收盘价状态概率向量: $\pi(3) = \pi(2)P = (0.4117, 0.0453, 0.5430)$ 。

由第 10 日收盘价状态概率向量 $\pi(1)$ 的结果可知, 收盘价落在 i_3 状态的概率最大, 为 0.579, 其该天的实际价格为 14.55, 的确处于下降状态。同理, 在第 11 日收盘价状态概率向量 $\pi(2)$ 中, 收盘价处于 i_3 状态的概率最大, 而事实上该天的实际价格为 14.59, 处于上身状态, 与预测结果不服, 因此投资者在实际决策中可以结合当天的实事来判断其所处的真实状态。

由于上述递推法计算量过大, 在此不一一列出, 但随着交易日的增加, 在天数足够多时, 状态概率会趋向于一个与初始值无关的值并稳定下来, 因此利用平稳分布与极限分布, 通过在稳定条件下进行分析决策。

按照马尔可夫链的稳定条件:

$$\begin{cases} \pi P = \pi \\ \sum_{i=1}^n x_i = 1, \pi = (x_1, x_2, x_3) \end{cases}$$

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

将该股的收盘价格状态向量 $\pi = (x_1, x_2, x_3)$, 其中 x_1 代表上升, x_2 代表持平, x_3 代表下降, 将状态转移矩阵 P 带入上述式中:

$$\begin{cases} (x_1, x_2, x_3) \begin{bmatrix} 0.3684 & 0.0526 & 0.5790 \\ 0 & 0 & 1 \\ 0.4783 & 0.0434 & 0.4783 \end{bmatrix} = (x_1, x_2, x_3) \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

由矩阵乘积得方程组:

$$\begin{cases} 0.3684x_1 + 0.4783x_3 = x_1 \\ 0.0526x_1 + 0.0434x_3 = x_2 \\ 0.5790x_1 + x_2 + 0.4783x_3 = x_3 \\ x_1 + x_2 + x_3 = 1 \end{cases}$$

解方程组得: $x_1 \approx 0.4115$ 、 $x_2 \approx 0.0452$ 、 $x_3 \approx 0.5433$ 。

可以看出, 由稳定状态计算出的收盘价格状态概率值与 $\pi(1)$ 、 $\pi(2)$ 的递推公式所推导的结论一致, 即该股最终以 41.55% 的可能性处于上升状态, 以 4.52% 的可能性处于持平状态, 以 54.33% 的可能性处于下降状态, 因此从短期来看, 投资者应该在所预测的 9 日前, 抛出这只股票相对比较安全。

4.2. 以各日收盘价格状态区间为对象进行预测

仍以平安银行(SZ000001)的收盘价格资料为例, 将 45 个交易日的收盘价格区间由低到高依次划分

为4个价格状态区间,每个区间长度为0.75个单位,得到区间状态为 $S_1 = [13.60, 14.35)$ 、 $S_2 = [14.35, 15.10)$ 、 $S_3 = [15.10, 15.85)$ 、 $S_4 = [15.85, 16.60)$,对数据进行统计可得下表3:

Table 3. Classified statistical table of stock price interval

表 3. 股价区间分类统计表

| 价格状态 | S_1 | S_2 | S_3 | S_4 |
|------|----------------|----------------|----------------|----------------|
| 价格区间 | [13.60, 14.35) | [14.35, 15.10) | [15.10, 15.85) | [15.85, 16.60) |
| 频数 | 4 | 15 | 15 | 11 |

整理数据得出不同状态区间的一步转移频数表4:

Table 4. Frequency table of one-step transfer

表 4. 一步转移频数表

| 状态 | S_1 | S_2 | S_3 | S_4 | Σ |
|-------|-------|-------|-------|-------|----------|
| S_1 | 2 | 2 | 0 | 0 | 4 |
| S_2 | 2 | 9 | 4 | 0 | 15 |
| S_3 | 0 | 4 | 7 | 4 | 15 |
| S_4 | 0 | 1 | 4 | 6 | 11 |

由表4统计可知,状态转移概率分别为:

$$P_{11} = P_{12} = 1/2, P_{13} = P_{14} = 0, P_{21} = 2/15, P_{22} = 9/15, P_{23} = 4/15, \dots$$

得状态转移矩阵:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{2}{15} & \frac{9}{15} & \frac{4}{15} & 0 \\ 0 & \frac{4}{15} & \frac{7}{15} & \frac{4}{15} \\ 0 & \frac{1}{11} & \frac{4}{11} & \frac{6}{11} \end{bmatrix}$$

根据原始资料可知,第45个交易日收盘价格为14.55,属于 S_2 状态,而后续无转移状态,所以假设其初始状态向量 $\pi(0) = (0, 1, 0, 0)$ 。则由马尔可夫链预测第45个交易日收盘价格状态概率向量:

$$\pi(1) = \pi(0)P = (0, 1, 0, 0) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{2}{15} & \frac{9}{15} & \frac{4}{15} & 0 \\ 0 & \frac{4}{15} & \frac{7}{15} & \frac{4}{15} \\ 0 & \frac{1}{11} & \frac{4}{11} & \frac{6}{11} \end{bmatrix} = \left(\frac{2}{15}, \frac{9}{15}, \frac{4}{15}, 0 \right)$$

可知,第45个交易日在 S_2 状态的概率最大,即5月09日的股票价格大概在14.35到15.10之间,

而事实上第 45 个交易日收盘价格为 14.55, 的确处于所预测的区间, 因此所预测的结果是正确的。如果利用稳定条件下的平稳分布, 所预测的最终结果与上述递推方式结果相同。

5. 结论

运用马尔可夫链建立的股价分析模型对了解股价运动以及预测股价走势有很大的帮助, 通过数据直观地了解一只股票的行情, 避免了很多因盲目而造成的损失。本文利用马尔可夫链, 在连续观察股票收盘价格的基础上, 以划分价格状态和价格区间两种方式为对象建立马尔可夫模型, 在对模型中股票的“马氏性”分析和股价分析的基础上, 对股票的价格走势进行了分析和预测, 通过将部分预测结果与实际资料对比发现, 两种划分方式下建立的马尔可夫模型所预测的结果均能与实际结果较好吻合, 很好地说明了在股价预测上, 该模型运用的可行性。但股市波动是一个复杂的非线性系统, 股票价格的变化受到多种因素的影响, 其包括利率、汇率、通货膨胀率、行业前景、预期心理、国家政策以及突发事件如新冠肺炎疫情等各个方面[5]。而股票市场环境的微小变化就可能引起股票价格的大幅度波动, 因此不存在绝对的变化规律, 所以没有任何一种预测方法可以准确地预测出股价每日的变化, 运用马尔可夫预测法也不例外。

此外, 本文对于初始状态向量与状态转移概率的确定, 是基于其在分析过程中保持稳定不变, 如果忽视此条件, 在时间上进行较大跨度的外推, 可能会产生较大的误差, 解决这一问题需要结合实际情况对初始状态向量与状态转移概率进行不断调整, 从而提高预测的可信度。由于长期下状态转移概率将发生变化, 使得大样本不一定能带来预测结果的高准确率, 因此进行预测时, 要注意针对不同的股票选取合适的样本容量。同时要注意的是, 模型所预测的结果只表明预测对象未来将以某一概率趋向于某种状态, 而不是绝对地处于这种状态, 因此预测结果仅提供一个进行决策参考的依据, 投资者需结合实际情况做出正确抉择, 并不能一味地信赖或依靠该模型, 股市有风险, 投资需谨慎!

基金项目

安徽省教学研究项目资助(2016jyxm0136)。

参考文献

- [1] 刘致谦. 基于马尔可夫链的股票价格波动预测[J]. 合作经济与科技, 2021(24): 64-67. <https://doi.org/10.3969/j.issn.1672-190X.2021.24.024>
- [2] 张波, 张景肖. 应用随机过程[M]. 北京: 清华大学出版社, 2004.
- [3] 夏莉, 黄正洪. 马尔可夫链在股票价格预测中的应用[J]. 商业研究, 2003(10): 62-65. <https://doi.org/10.3969/j.issn.1001-148X.2003.10.023>
- [4] 周旋. 马尔可夫链在股票中的运用[J]. 中外企业家, 2013(13): 105. <https://doi.org/10.3969/j.issn.1000-8772.2013.13.041>
- [5] 台文志. 利用马尔可夫链模型预测股票市场的近期走势[J]. 西南民族大学学报(自然科学版), 2008, 34(3): 477-481. <https://doi.org/10.3969/j.issn.1003-2843.2008.03.019>