1.3 本文的研究工作和主要结构

1.3.1 研究工作

本文应用随机过程中的 Markov 链及 HMM 模型相关理论研究中国石油股票背景及其影响，在此基础上建立股票价格预测模型，利用 MATLAB 实现股票价格变化趋势的预测。研究工作主要从以下三个部分开展：

第一部分，分析股票价格的历史数据，利用 Markov 预测方法建立股票价格预测模型来预测股票价格的未来走势。

第二部分，基于 Markov 预测的缺陷，进一步利用 HMM 模型对该问题进行了更为细致地思考，并利用该模型进一步预测。

第三部分，将理论运用于实证研究，对 Markov 链模型与 HMM 模型的预测结果进行对比分析，并给出结论。

1.3.2 主要结构

本文共分为五个部分：

第一章：绪论。介绍了本文相关课题的研究背景和意义，国内外研究现状，以及本文的主要学术研究方向和理论研究结构。

第二章：相关理论和基础知识。介绍了 Markov 链模型的相关概念与 Markov 预测步骤，以及 HMM 模型的简介，算法和预测步骤。

第三章：Markov 预测法在中国石油股票价格预测中的应用。首先将中国石油股票的数据进行预处理，之后进行模型的假设、马氏性检验与建立模型，同时对模型进行 MATLAB 实现，最后对模型进行评价。

第四章：HMM 模型在中国石油股票价格中的应用。首先对数据进行预处理，其次，建立并检验模型，同时利用 MATLAB 实现了该模型，最后对 Markov 链模型与 HMM 预测模型对比分析。

第五章：结论。总结全文研究成果，找出研究不足和未来分析研究的发展前景。

本文的技术流程图如下图 1-1 所示:

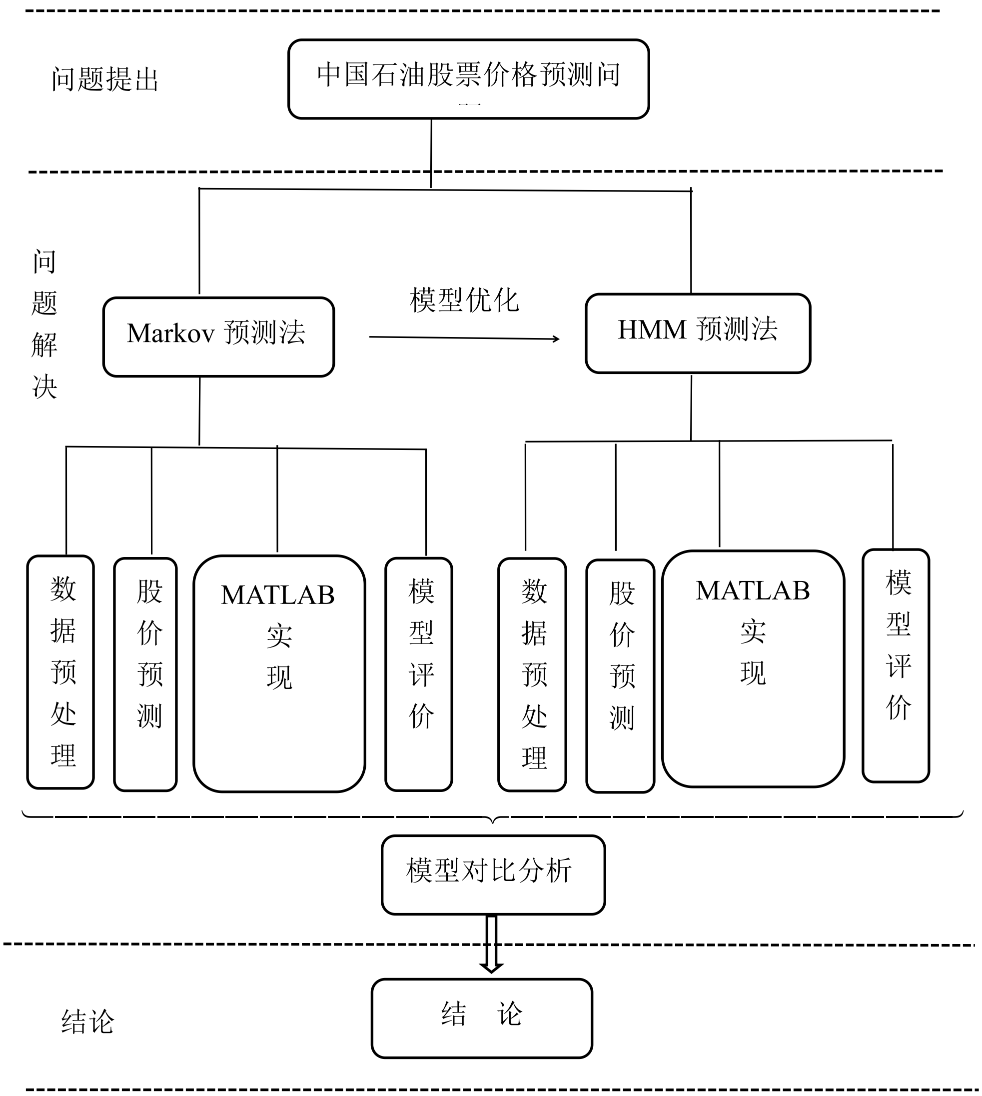

图 1-1 技术流程图

2 相关理论及基础知识

2.1 Markov 预测法的理论概述

Markov 过程, 因俄罗斯著名的数学 Andre Markov 而得名。具有 Markov 属性的离散随机过程，即 Markov 链。Markov 链在很多应用中扮演着重要的作用。例如，Google 的网页分类算法由 Markov 链确定。Markov 链是一种特殊的随机过程。

2.1.1 Markov 链的基本概念

（1）Markov 链的定义

定义 2. 1 设 $\{X(t), t\in T\}$，其中，时间 $T=\{0,1,\cdots\}$，状态空间 $I=\{0,1,2,\cdots\}$，若对任意时刻 n 及其任意状态 $i_0, i_1, \cdots, i_{n-1}, i, j$，有

$$P\{X(n+1)=j \mid X(n)=i, X(n-1)=i_{n-1}, \cdots, X(1)=i_1, X(0)\}=P\{X(n+1)=j \mid X(n)=i\}$$

则称 $\{X(t), t\in T\}$ 为 Markov 链。

（2）马氏性

定义 2. 2 如果知道现在的状态（$t=\mu$ 时 X_μ 的状态），未来（$X_\nu(\nu>\mu)$）的状态不取决于过去（$X_w(w<\mu)$），那么就说这个过程 $\{X_n\}$ 具有马氏性，也可称之为“无后效性”。马氏性表明，当知道“现在”时，“未来”和“过去”是独立的。

R. A. Howard 为提供了对 Markov 链风景如画般的描述，就像一只青蛙在一组睡莲上跳跃。这只青蛙开始在一片叶子上，然后以相应的转移概率从一片睡莲叶子跳到另一片叶子上。

易见，上述 Markov 链具有马氏性。

（3）状态转移概率

定义 2. 3 在一个事件发生的过程中，从一个状态过渡到下一个状态的可能性称为状态转移概率。

研究一随机过程 $\{X_n, n=0,1,2\cdots\}$，它只需要有限的数量或多个值，这个过程的所有可能的值被称为状态空间，通常记录在 E，假设 $E=\{0,1,2\cdots\}$。

若 $X_n=i$，就说“过程在时刻 n 处于状态 i”，假设每当过程处于状态 i，下一个时刻将处于状态 j 的概率是固定的 p_{ij}，那么在任意时刻 n，

$$P(X_{n+1}=j \mid X_n=i)=p_{ij} \tag{2-1}$$

p_{ij} 代表处于状态 i 过程于下一步转移到状态 j 的概率。由于概率为非负且过程必须转移到某一状态，状态转移概率满足以下条件：

① $0\le p_{ij}\le 1, i, j=0,1,\cdots,n$

② $\sum_{j=1}^{n} p_{ij}=1$

（4）状态转移概率矩阵

定义 2. 4 以 P 记一步转移矩阵 p_{ij} 的矩阵，

$$P=\begin{bmatrix} p_{00}\cdots p_{0j}\cdots \\ p_{10}\cdots p_{1j}\cdots \\ \vdots \quad\quad \vdots \\ p_{i0}\cdots p_{ii}\cdots \\ \vdots \quad\quad \vdots\ \cdots \end{bmatrix}$$

P 就是所谓的一步转移概率矩阵，简称转移矩阵。矩阵 P 的第 $i+1$ 行就是给定 $X_n=i$ 时 X_{n+1} 的条件概率分布。

2.1.2 Markov 链的状态向量及平稳分布

用 $\pi_j(k)$ 表示事件在第 k 个时刻（时期）处于状态 j 的概率。

显然，$\sum_{i=1}^{n}\pi_j(k)=1$。由 Bayes 条件概率公式和 Markov 过程的无后效性有

$$\pi_j(k)=\sum_{i=1}^{n}\pi_j(k-1)p_{ij}\quad i,j=0,1,\cdots,n \tag{2-2}$$

记 $\pi(k)=[\pi_1(k),\cdots,\pi_n(k)]$ 为第 k 个时间（周期）的状态概率向量。根据上述公式,给出了状态概率向量的递推计算公式:

$$\pi(k)=\pi(k-1)P=\cdots=\pi(0)P^k \tag{2-3}$$

其中，$\pi(0)$ 为初始状态概率向量。

因此，如果知道一个事件在某个特定时间（周期）的状态 π（s_0），那么使用概率矩阵 P 和状态转移的递推公式，可以计算出在第 k 个转移周期之后的第 s_0+k 个特定时间（周期）内几个可能状态的概率 $\pi(s_0+k)$。故可以得到第 s_0+k 个特定时间（时段）事件状态转换为预测的概率。

定义 2. 5 设 Markov 链 $\{X(t),t\in T\}$ 中一个具有转移概率矩阵为 P，若存在某一个概率分布 $\{\pi_j,j\geq 0\}$，其满足

$$\pi_j=\sum_{i=0}^{\infty}\pi_i p_{ij},j=0,1,2,\cdots$$

则称 $\{\pi_j,j>0\}$ 为该 Markov 链的平稳分布。

由定义知，若 $\pi=\{\pi_0,\pi_1,\cdots\}$，则 $\pi=\pi P$。

一般来说，平稳分布可能不存在，如果平稳分布是存在的，并且是唯一的，那么它是以下代数方程的唯一的非负解：$\begin{cases}\pi=\pi P\\ \sum_{i=1}^{n}\pi_i=1\end{cases}$

定理 2. 1 如果状态空间 E（不妨 $E=i_0,i_1,\cdots,i_n$）和 Markov 链 $\{X_n,n=1,2,\cdots\}$ 是一个有限集，其转移概率矩阵是互达的，则它具有唯一的平稳分布；

$$\pi=(\pi_1,\cdots,\pi_n)$$

且

$$\lim_{n\to\infty}P^{(n)}=\lim_{n\to\infty}P^n=\begin{pmatrix}\pi_1 & \pi_2 & \cdots & \pi_n\\ \pi_1 & \pi_2 & \cdots & \pi_n\\ \cdots & \cdots & \cdots & \cdots\\ \pi_1 & \pi_2 & \cdots & \pi_n\end{pmatrix}$$

2.1.3 Markov 预测法

Markov 预测法是一种针对某一事件，预测该事件将来会发生某种概率的方法。它的主要原理是根据具体事例的具体情况，以 Markov 链为理论基础，对这个事件在未来某一段时间内可能发生的概率进行预测。

使用 Markov 链预测事件发生概率是有一定条件的，并不是在什么情况下都可行的，稳定的状态转移概率矩阵是它的基本要求。要保证概率矩阵的稳定性，预测是就必须要用相当的数据；所用数据越多，越能够反映某一事件的一般性，则预测的结果就越稳定。

（1）Markov 预测前提

运用 Markov 预测法研究实际问题时，预测期内适用于对象的要求：

① 保证事件的转移概率的一致性，使事件的状态转移的概率在预测时段内的每个周期内都是不变的，即要保证从一个步骤进行到下一个步骤的转移概率变。

② 在预测期内，要保证预测的状态数一致。

③ 状态转换只依赖于前一时期，与前一时期之前的状态无关。

（2）Markov 预测步骤

运用 Markov 预测法解决问题的步骤如下：

① 确定预测对象的状态和状态空间

$$E=[i_0,i_1,\cdots,i_n]$$

② 确定转移概率 p_{ij} 和状态转移概率矩阵 P；

③ 计算初始状态分布向量

$$\pi(0)=[\pi_1(0),\cdots,\pi_n(0)]$$

④ 建立模型进行预测，计算 k 时刻各状态的状态向量

$$\pi(k)=[\pi_1(k),\cdots,\pi_n(k)]$$

⑤ 解方程组

$$\begin{cases} \pi = \pi P \\ \sum_{i=1}^{n} \pi_i = 1 \end{cases}$$

得终极状态π=（$\pi_1, \pi_2, \cdots, \pi_n$）。

3 Markov 链模型在中国石油股价预测中的实证研究

上一章研究了 Markov 链模型，因此依次进行了具体的研究。实证研究主要通过观察中石油股票价格的走势来发现中石油股票价格的整体变化，进而划分状态空间。针对这三种变化，研究每日收盘价变化，将增长率按照分类划分为不同区间，确定每个区间的状态数，并进行频数统计，为未来预测奠定基础。建立 Markov 链模型。通过频率统计，计算一步转移概率矩阵，验证其 Markov 性质，预测未来股票价格走势。

3.1 数据预处理

3.1.1 数据的选取

本文选取中国石油股票 2007 年 11 月 5 日至 2021 年 2 月 26 日，一共 3230 个交易日的收盘价、涨跌幅度、成交量等指标的数据为例。这些交易日中收盘价的最大值为 40. 4300，最小值为 4. 0500，平均值为 9. 6832。所用数据来源于大智慧软件。

3.1.2 股价日涨跌幅趋势分析

与前一交易日收盘价相比，各交易日收盘价呈现上升、持平、下降三种走势，并对三种状态进行了分析和预测。利用 MATLAB 将各日的收盘价绘制成图，如图 3-1 所示。横轴表示股票价格的时间序列，纵轴表示股票价格。从图中可以看出中国石油股票价格走势先下跌后上涨然后下跌。

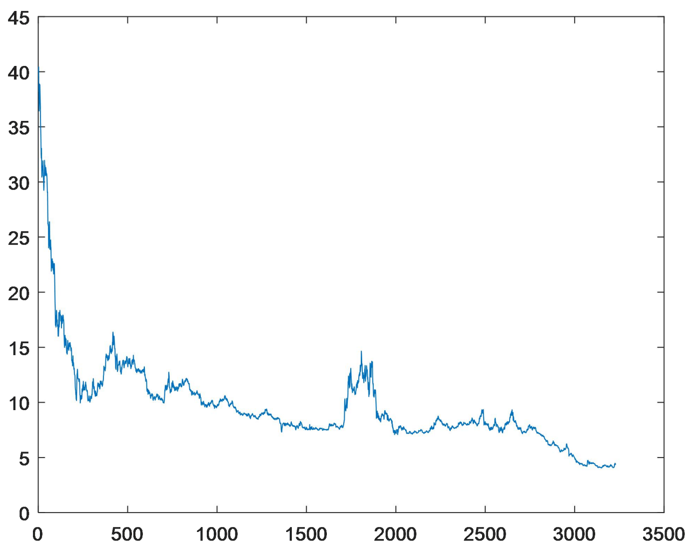

图 3-1 中国石油股票价格走势图

3.2 预测模型应用及其 MATLAB 实现

3.2.1 一步转移概率矩阵

由于股票价格无时无刻都在波动，当在某一小范围内波动时可以把它看作没有变化。如果股票价格高于前一交易日，则状态为“涨”，以 1 表示；如果股票价格与上一交易日持平，则状态为“持平”，用 2 表示；如果股票价格低于前一交易日，则状态为“下跌”，用 2 表示。选取± 0.06为允许的波动范围。波动幅在$(0.06,+\infty)$记为状态 1，波动幅度在$[-0.06,+0.06]$内记为状态 2，波动幅在$(-\infty,-0.06)$记为状态 3。部分中国石油股票价格的状态统计如表 3-1 所示。

表 3-1 中国石油股票部分交易日状态表

日期	状态	日期	状态	日期	状态
2007/11/5	2	2007/11/19	3	2007/12/3	3
2007/11/6	3	2007/11/20	3	2007/12/4	1
2007/11/7	1	2007/11/21	3	2007/12/5	1
2007/11/8	3	2007/11/22	3	2007/12/6	3
2007/11/9	2	2007/11/23	3	2007/12/7	1
2007/11/12	3	2007/11/26	3	2007/12/10	1
2007/11/13	1	2007/11/27	3	2007/12/11	1
2007/11/14	1	2007/11/28	3	2007/12/12	3
2007/11/15	3	2007/11/29	1	2007/12/13	3
2007/11/16	1	2007/11/30	3	2007/12/14	2

转移概率矩阵主要是通过 MATLAB 软件，根据状态划分的结果，统计出一步状态转移的频率，用概率代替频率的。具体的 MATLAB 程序如下：

```
datas = S4;
E = datas(:,2)';
for i = 1:3
   for j = 1:3
         f(i,j) = length(findstr([i j],E));
   end
end
f %输出状态转移矩阵
fs = (sum(f'))';
```

fs %输出状态转移概率矩阵

for i = 1:3

  p(i,:) = f(i,:)/fs(i);

end

p %输出状态转移概率矩阵

得到的 MATLAB 运行结果如图 3-2 所示：

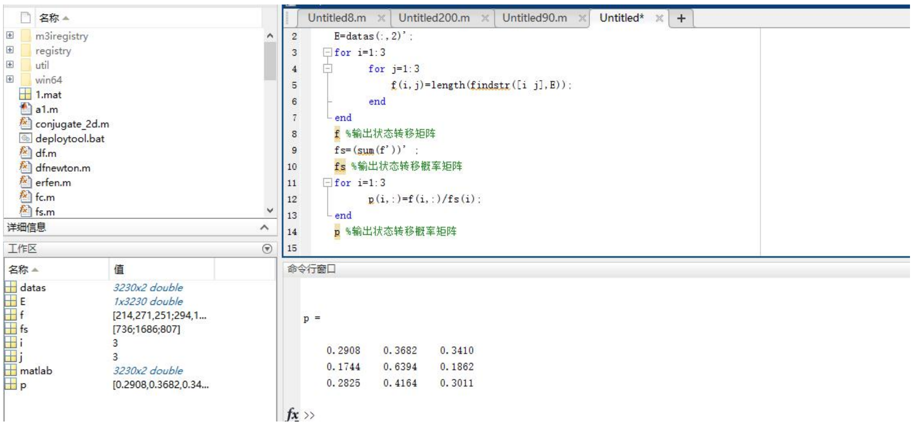

图 3-2 状态频数转移矩阵的 MATLAB 运行图

通过 MATLAB 实现可知，上升状态为 724 天，持平状态为 1686 天，下降状态为 808 天，经过状态划分后，用 MATLAB 统计出从一个状态到另一个状态的频率，并计算出从一个阶段到另一个阶段的概率矩阵。

状态频数转移的结果如表 3-2 所示。

表 3-2 频数统计表

后期 前期	1	2	3
1	214	271	251
2	294	1078	314
3	228	336	243

故可得各状态频数转移矩阵为：

$$f=\begin{bmatrix}214 & 271 & 251\\ 294 & 1078 & 314\\ 228 & 336 & 243\end{bmatrix}$$

由状态频数转移矩阵计算得到的一步状态转移概率矩阵为：

$$P=\begin{bmatrix}0.2908 & 0.3682 & 0.3410\\ 0.1744 & 0.6394 & 0.1862\\ 0.2825 & 0.4164 & 0.3011\end{bmatrix}$$

从一步转移概率矩阵可以看出，所有状态只需要一步就可以回到自身的状态，因此该链是非周期性的；所有状态离任何状态也只有一步之遥，因此该链是不可约的。该链只有三个状态，所以它是有限的。综上所述，该链具有唯一的平稳分布。

3.2.2 马氏性检验

利用 Markov 预测方法进行预测，需要对测试序列进行马氏性检验。

记 f_{ij} 为状态 i 到状态 j 的频数，p_{ij} 为状态 i 到状态 j 的概率，$p_{ij}=f_{ij}\Big/\sum_{j=1}^{m}f_{ij}$，$p_{ij}^{(n)}$ 表示从状态 i 到状态 j 的 n 步转移概率。将转换概率矩阵中的列 j 之和除以各行列之和，以获得边际概率 $p_{\bullet j}$，即 $p_{\bullet j}=\sum_{i=1}^{m}f_{ij}\Big/\sum_{i=1}^{m}\sum_{j=1}^{m}f_{ij}$，则统计量 $\chi^2=2\sum_{i=1}^{m}\sum_{j=1}^{m}f_{ij}\left|log\frac{p_{ij}}{p_{\bullet j}}\right|$，使以自由度为 $(m-1)^2$ 的 χ^2 为极限分布。

给定显著性水平 α，若 $\chi^2>\chi_{\alpha}^{2}(m-1)^2$，则认为该序列具有马氏性，否则不能用 Markov 链模型来预测股票价格。

由 $p_{\bullet j}$ 算得的边际概率如表 3-3 所示。

表 3-3 边际概率表

状态	1	2	3
$p_{\bullet j}$	736/3230	1686/3230	808/3230

计算的极限分布值如表 3-4 所示。

表 3-4 极限分布计算表

状态	$f_{i1}\left\|log\frac{p_{i1}}{p_{\bullet 1}}\right\|$	$f_{i2}\left\|log\frac{p_{i2}}{p_{\bullet 2}}\right\|$	$f_{i3}\left\|log\frac{p_{i3}}{p_{\bullet 3}}\right\|$	合计
1	52. 192	94. 5806	77. 7604	224. 533
2	78. 6146	218. 7243	92. 7112	390. 0501
3	49. 0042	75. 9312	45. 0431	169. 9785
合计	179. 8108	389. 2361	215. 5147	784. 5616

从表 3-4 中可以看出，在给定显著性水平 $\alpha = 0.05$ 的情况下，$\chi_{0.05}^2(3-1)^2 = 9.448$，$\chi^2 = 784.5616 > \chi_{\alpha}^2(m-1)^2$，故该股票序列满足马氏性。

3.2.3 建立模型进行预测

本文将 2021 年 3 月 1 日至 2021 年 3 月 26 日的数据在作为预测的样本，故将 2021 年 2 月 26 日的股票价格状态作为初始分布向量记为：$\pi(0) = (0,1,0)$，以此来判断未来股票价格的涨跌情况，从而推断出未来交易日的收盘价。

建立模型进行预测，计算 k 时刻各状态的状态向量为

$$\pi(k) = [\pi_1(k), \cdots, \pi_n(k)],\ k = 1, \cdots, 20$$

利用状态转移概率矩阵和递推公式，预测了 2021 年 3 月 1 日至 2021 年 3 月 26 日几个可能状态的概率，具体的 MATLAB 程序如下：

```
S{1} = [0 1 0];
for i = 1:20
  S{i+1} = S{i}*p;
end
S{2:end}
```

通过 MTALBA 得到的各状态的概率结果为：

```
ans = 0.1744    0.6394    0.1862
ans = 0.2148    0.5506    0.2346
ans = 0.2248    0.5288    0.2464
ans = 0.2272    0.5235    0.2493
ans = 0.2278    0.5222    0.2500
ans = 0.2279    0.5218    0.2502
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
ans = 0.2280    0.5217    0.2503
```

ans = 0.2280　　0.5217　　0.2503

ans = 0.2280　　0.5217　　0.2503

ans = 0.2280　　0.5217　　0.2503

ans = 0.2280　　0.5217　　0.2503

由上可知 2021 年 3 月 1 日的状态概率向量为 $\pi(1) = (0.1744, 0.6394, 0.1862)$，$max = \{(0.1744, 0.6394, 0.1862)\} = 0.6394$，所以收盘价处于第二状态的可能性最大，2021 年 3 月 1 日的收盘价为 4. 31，从数值上看处于第 2 状态，与预测结果相符。其他交易日的预测结果以此类推，预测结果如表 3-5 所示。

表 3-5 预测结果表

日期	状态向量			最大值	预测结果	实际结果	
	1	2	3		状态	状态	预测值
2021/3/1	0.1744	0.6394	0.1862	0.6394	2	2	4.37
2021/3/2	0.2148	0.5506	0.2346	0.5506	2	3	4.27
2021/3/3	0.2248	0.5288	0.2464	0.5288	2	2	4.31
2021/3/4	0.2272	0.5235	0.2493	0.5235	2	2	4.33
2021/3/5	0.2278	0.5222	0.2500	0.5222	2	1	4.42
2021/3/8	0.2279	0.5218	0.2502	0.5218	2	1	4.51
2021/3/9	0.2280	0.5217	0.2503	0.5217	2	2	4.47
2021/3/10	0.2280	0.5217	0.2503	0.5217	2	3	4.39
2021/3/11	0.2280	0.5217	0.2503	0.5217	2	2	4.44
2021/3/12	0.2280	0.5217	0.2503	0.5217	2	2	4.45
2021/3/15	0.2280	0.5217	0.2503	0.5217	2	2	4.51
2021/3/16	0.2280	0.5217	0.2503	0.5217	2	2	4.47
2021/3/17	0.2280	0.5217	0.2503	0.5217	2	2	4.37
2021/3/18	0.2280	0.5217	0.2503	0.5217	2	2	4.27
2021/3/19	0.2280	0.5217	0.2503	0.5217	2	3	4.31
2021/3/22	0.2280	0.5217	0.2503	0.5217	2	2	4.33
2021/3/23	0.2280	0.5217	0.2503	0.5217	2	2	4.42
2021/3/24	0.2280	0.5217	0.2503	0.5217	2	2	4.51
2021/3/25	0.2280	0.5217	0.2503	0.5217	2	2	4.47
2021/3/26	0.2880	0.5217	0.2503	0.5217	2	2	4.39

3.3 模型评价

由表 3-5 可知，预测结果中只有 2021 年 3 月 2 日、5 日、8 日、10 日、19 日预测结果如实际相悖，即在预测的 20 日中有 15 日的预测结果如实际相符，只有 5 日预测错误，正确率为 75%。

Markov 链模型是一种预测未来变化过程的技术。它主要利用 Markov 基本原理和方法分析变化规律；它的问题的实质是在概率条件下寻找期望值；使用它的关键是得到事物的原始向量和状态转移概率矩阵，但它有一定的局限性。首先，它必须基于某些很难在显示市场上很难做出的假设，由于在短期内这些假设很容易实现，所以结果更准确；其次，Markov 链模型忽略了事物发展的次要因素，这使得预测结果不太准确。总之，利用 Markov 链模型预测股票价格还是有很强的应用价值