DOI:10.14018/j.cnki.cn13-1085/n.2010.24.233

灰色-马尔可夫链模型在股市预测中的应用

Gray-The Application of Markov Chain Model in Stock Forecasting

简艳群 Jian Yanqun

（华北电力大学，保定 071003）

（North China Electric Power University，Baoding 071003，China）

摘要：用 GM（1，1）预测具有良好的精确性和规律性，但对于随机波动性较大的股市行业，它的预测精度比较低，而马尔可夫模型可以克服波动性较大的局限性，弥补灰色模型的不足，因此将两者结合起来对股市进行预测将能提高预测的精度。本文依据上交所 20 个月末收盘指数预测后四个月的月末收盘指数范围，实证分析表明灰色马尔可夫链模型在股市预测中应用的可行性

Abstract: Using GM（1，1）model to predict has great accuracy and regularity.But for the random high waving stock market，The accuracy is low. But Markov model can overcome the defection of high waving and make up the shortage of GM （1，1） model. So combining GM （1，1） with Markov chain model to predict stock market can improve the precision of prediction. Based on the indexes ended in the twenty months in Shanghai Stock Exchange，the range of the index in the end of next four months was predicted. Empirical analysis shows that GM - Markov chain model is a feasible tool to predict the stock market.

关键词：灰色预测模型；马尔可夫模型；月末上证收盘指数；预测

Key word: gray prediction；markov model；the index of shanghai stock exchange close in the end of month；predict

中图分类号：F22　　文献标识码：A　　文章编号：1006-4311（2010）24-0255-02

0 引言

在股票市场中，股票价格是一个基本特征量，但是它总受政治、经济等各方面的影响，具体的影响因素的程度和信息是不完全的，所以我们可以把股市当成一个灰色系统来处理。灰色 GM（1，1）预测模型是灰色系统理论的重要组成部分，主要适用于时间短、数据资料少、波动性不大的预测问题，且只需很少的几个数据即可建立模型进行预测，可以很好地解决由于数据少而导致的精确度不高的问题，但由于灰色 GM（1，1）预测模型的预测曲线是一条较平滑的单调曲线，对波动性较大的股票市场中的数据列拟合较差，预测度较低。同时马尔可夫链比较适合随机波动性较大的预测问题，但是马尔可夫链要求状态无后效性，且要具有平稳过程等特点。如果灰色 GM（1，1）模型对数据进行拟合，找出其变化趋势，则可以弥补马尔可夫预测的局限性，而在灰色预测基础上进行马尔可夫预测，又可弥补灰色预测对随机波动性较大的数据序列准确度低的不足，所以可将二者结合起来。形成灰色马尔可夫预测模型，对随机波动性较大的数据列进行预测，大大提高随机波动性较大数据列的预测精度，为随机波动性较大的对象的预测提供一种新的方法。

1 灰色-马尔可夫链预测模型

1.1 灰色模型 GM（1，1）

1.1.1 给定原属数据列，记作 $x^{(0)}=[x^{(0)}(k)|k=1,2,\cdots,n]$

1.1.2 对原始数据进行一次累加，生成新的数据序列记 $x^{(1)}=[x^{(1)}(k)|k=1,2,\cdots,n]=[x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)]$

其中 $x^{(1)}$ 与 $x^{(0)}$ 之间满足下述关系 $x^{(1)}(k)=\sum_{i=1}^{k}[x^{(0)}(i)]$

1.1.3 构造矩阵 B 和常数矩阵 Y_n。

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=u \quad Y_n=BA$$

$$B=\begin{bmatrix}-\frac{1}{2}[x^{(1)}(1)+x^{(1)}(2)] & 1\\ -\frac{1}{2}[x^{(1)}(2)+x^{(1)}(3)] & 1\\ \vdots & \vdots\\ -\frac{1}{2}[x^{(1)}(n-1)+x^{(1)}(n)] & 1\end{bmatrix} \quad Y_n=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix} \quad A=\begin{bmatrix}a\\ u\end{bmatrix}$$

1.1.4 用最小二乘法求解 a，u。

$$A=(B^TB)^{-1}B^TY_n=\begin{bmatrix}\hat{a}\\ \hat{u}\end{bmatrix}$$

1.1.5 该模型的时间响应方程为：

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k}+\frac{\hat{u}}{\hat{a}} \quad (k=0,1,2\cdots)$$ 再经累减还原得

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$$

$$=(1-e^{\hat{a}})\left(\hat{x}^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right)e^{-\hat{a}k} \quad (k=0,1,2,\cdots)$$

1.2 马尔可夫链预测模型

1.2.1 状态划分。

①根据 GM（1，1）模型求出原始数据序列的拟合值 $x(i)$；

②求出残差 $\Delta k=x(i)-\hat{x}(i)$；

③残差的相对值为 $\varepsilon(i)=\frac{x(i)-\hat{x}(i)}{\hat{x}(i)}\times100\%$；

④为了使每一状态的数据相差不多，将 $\varepsilon(i)$ 的值从小到大排列，根据用户的需要和数据的多少，将状态分为自己想要的数。

1.2.2 马尔可夫链。

①马尔可夫链的概念。

定义 1 设随机过程$\{X(t),t\in T\}$的状态空间 S 为 R 中的可列集。如果对 T 中任意 n 个参数 $t_1<t_2<\cdots<t_n$，以及使 $P\{X(t_n)=i_n|X(t_{n-1})=i_{n-1},\cdots,X(t_1)=i_1\}=P\{X(t_n)=i_n|X(t_{n-1})=i_{n-1}$ (1)

则称$\{X(t),t\in T\}$为马尔可夫链。

定义 2 若马尔可夫过程的状态空间 S 为 R 中的可列集，时间参数集 T 为可列离散集，则称 $P=\{X_{m+1}=j|X_m=i,X_{m-1}=i_{m-1},\cdots,X_0=i_0\}=P\{X_{m+1}=j|X_m=i\}$ (2)

定义 3 若 $P\{X_{m+1}=j|X_m=i\}\equiv P\{X_1=j|X_0=i\}\equiv P(m,1)\equiv P_{ij}$ 即从 i 状态转移到 j 状态的概率与 m 无关，称这类马尔可夫链为齐次马尔可夫链，P_{ij} 表示由状态 i 经过一步转移到状态 j 的概率，称为一步概率转移，以 P_{ij} 为元素的矩阵 $P=(P_{ij})$ 称为状态一步转移概率矩阵，其形式为：

$$P=\begin{bmatrix}P_{11} & \cdots & P_{1k}\\ \vdots & & \vdots\\ P_{k1} & \cdots & P_{kk}\end{bmatrix}$$

其中 $P_{ij}\geq0$，$(i,j\in s)$；$\sum_{j\in s}P_{ij}=1$，$(i\in s)$。

②转移概率的计算。

a.前面的状态划分，将残差的相对值分为若干状态，记为 E_1,E_2,E_3,E_4，残差相对值序列由状态 E_i 经 k 步转移到状态 E_j 的概率成为 n 步转移概率，记为 $p_{ij}(k)=\frac{m_{ij}(k)}{M_i}$，式中 $m_{ij}(k)$ 为状态 E_i 经 k 步转

作者简介：简艳群（1987-），女，浙江湖州人，本科学历，研究方向：工商管理。

移到状态 E_j 的次数，M_i 为状态 E_i 出现的次数，由于数据序列的最后状态的转向不明确，故计算 M_i 为时要去掉数据序列中最末的 k 个数据；

b.当 k=1 时，即为一步转移概率 P_{ij}，其矩阵形式可记为：

$$P=\begin{bmatrix}P_{11} & \cdots & P_{1k}\\ \vdots & & \vdots\\ P_{k1} & \cdots & P_{kk}\end{bmatrix}$$，其中 $P_{ij}\geq 0$，$(i,j\in s)$；$\sum_{j\in s}P_{ij}=1$，$(i\in s)$

c.$P(1)=PP(0)$，考察 $P(1)$ 中的 n 个值，若 $\max P_{kj}=P_{k1}$，则可以认为下一时刻系统最有可能由状态 E_k 转向状态 E_1，即下一时刻最有可能处于状态 E_1；

d.$P(n)=PP(n-1)$，同理上一步可知 n 时刻后系统最有可能所处状态。

e.计算预测区间及预测值。确定预测对象未来的转移状态转移以后，即确定了残差预测值的变动区间 E_j，我们用区间的中位数作为残差的预测值，即残差的预测值=（预测区间最大值+预测区间最小值）/2。则最终预测对象的预测值=（1+残差预测值）灰色预测值。

2 实例分析

由于收盘价是影响股票价格的最主要的因素，本文在上交所交易指数数据中选取从 2008 年 10 月到 2010 年 5 月 20 个月末收盘指数，其中时间序列的单位以月计。

步骤一：根据表 1 中已有的实际数据，用 Matlab 求出模型 GM（1，1）中的 a，u。a=0.0215，u=2160.4，则 $\hat{x}(t)=2160.4e^{0.0215t}$；

步骤二：由残差相对值公式 $\varepsilon(t)=\frac{x(t)-\hat{x}(t)}{\hat{x}(t)}\times 100\%$，求得残差相对值序列 $\varepsilon(t)$ 的范围为（-25.317%，23.1901%）；

步骤三：根据实际情况将残差相对值 $\varepsilon(t)$ 平均分为 4 个状态 a，b，c，d；其中 a=[13%，26%）；b=[0，13%）；c=[-13%，0）；d=[-26%，-13%）。

步骤四：残差相对值状态的一步转移矩阵为：

$$P=\begin{pmatrix}P_{11} & P_{12} & P_{13} & P_{14}\\ P_{21} & P_{22} & P_{23} & P_{24}\\ P_{31} & P_{32} & P_{33} & P_{34}\\ P_{41} & P_{42} & P_{43} & P_{44}\end{pmatrix}=\begin{pmatrix}0.5 & 0 & 0.5 & 0\\ 0.125 & 0.75 & 0.125 & 0\\ 0 & 0.4 & 0.2 & 0.4\\ 0 & 0 & 0.25 & 0.75\end{pmatrix}$$

用 2010 年 5 月上证指数残差相对值作为初始状态，认为初始分布 $I^{(0)}=(0\ \ 0\ \ 0\ \ 1)$。

则下个月的绝对分布为：

$$I^{(1)}=I^{(0)}\cdot p=(0\ \ 0\ \ 0\ \ 1)\cdot\begin{pmatrix}0.5 & 0 & 0.5 & 0\\ 0.125 & 0.75 & 0.125 & 0\\ 0 & 0.4 & 0.2 & 0.4\\ 0 & 0 & 0.25 & 0.75\end{pmatrix}=(0\ \ 0\ \ 0.25\ \ 0.75)$$

所以下个月预测对象最有可能处于状态 d，3，即 $\varepsilon(t)\in d=(13\%,26\%]$，又由 $\varepsilon(t)=\frac{x(t)-\hat{x}(t)}{\hat{x}(t)}\times 100\%$，得 $x(t)=\frac{\hat{x}(t)}{1-(t)}$，所以灰色马尔可夫预测区间 $x(21)\in(2691.19,3000.796)$，灰色马尔可夫预测值 $x(21)=2845.993$。

步骤五：同理再进行预测可以依照上面的步骤得到以后各个月份的预测区间及预测值。预测区间及预测值详见表 2。

由以上数据经灰色马尔可夫预测模型计算得，见表 2。

以上计算结果表明的灰色马尔可夫预测模型较单一的灰色预测模型和马尔可夫预测模型来说精度和准确性都有了明显的提高。

3 结论

灰色马尔可夫链预测模型是根据灰色模型和马尔可夫链模型思想建立的，同时建立在历史数据的统计分析基础上，因此历史数据越准确，精度越高。该模型用灰色 GM（1，1）模型预测曲线来反映上证指数的发展趋势，以此预测曲线为基准，再运用马尔可夫模型来寻找上证指数的转移状态，不仅考虑了数据序列中的演变规律，而且通过状态转移概率矩阵的变换提取数据中的随机响应，因而它将灰色模型和马尔可夫链模型的优点结合起来，克服各自的缺点提高预测精度，另外该模型的原理浅显易懂、计算过程不复杂，适用性比较强。该模型存在许多值得探讨的问题，它的预测精度与状态的划分有很大的关系，目前状态的划分没有统一的标准，所以还需要进一步的研究，另外该模型对波动性较大且有一定上升趋势的数据来说，可以取得比较好的预测效果，但并不是对任何数据都能适用。总之，灰色马尔可夫链预测模型有其应用价值，为投资者投资决策提供一定的理论依据。

表 1 上证指数（2008 年 10 月到 2010 年 5 月共 20 个月末收盘指数）

序号	月份	$x(t)$	$\hat{x}(t)$	$\varepsilon(t)$	状态
1	2008 年 10 月	1977.14	2160.4	-0.0926894	c
2	2008 年 11 月	1871.16	2207.2	-0.1795891	d
3	2008 年 12 月	1820.81	2255.1	-0.2385147	d
4	2009 年 1 月	1990.66	2304.1	-0.1574553	d
5	2009 年 2 月	2082.85	2354.1	-0.1302302	d
6	2009 年 3 月	2373.21	2405.2	-0.0134796	c
7	2009 年 4 月	2477.57	2457.3	0.0081814	b
8	2009 年 5 月	2632.93	2510.7	0.04642357	b
9	2009 年 6 月	2959.36	2565.2	0.13319096	a
10	2009 年 7 月	3412.06	2620.8	0.23190096	a
11	2009 年 8 月	2667.75	2677.7	-0.0037297	c
12	2009 年 9 月	2779.43	2735.8	0.01569746	b
13	2009 年 10 月	2995.85	2795.2	0.06697598	b
14	2009 年 11 月	3195.3	2855.8	0.1062498	b
15	2009 年 12 月	3277.14	2917.8	0.10965049	b
16	2010 年 1 月	2989.29	2981.1	0.00273978	b
17	2010 年 2 月	3051.94	3045.8	0.00201184	b
18	2010 年 3 月	3109.11	3111.9	-0.0008974	c
19	2010 年 4 月	2870.61	3179.4	-0.1075695	c
20	2010 年 5 月	2592.15	3248.4	-0.2531682	d

表 2 2010 年 6 月到 2010 年 9 月上证指数月末收盘预测区间及预测值

序号	月份	灰色预测模型预测值	灰色马尔可夫链预测模型预测的区间	灰色马尔可夫链预测模型预测值
21	2010 年 6 月	3390.9	（2691.19，3000.796）	2845.993
22	2010 年 7 月	3464.5	（2749.603，3065.929）	2907.766
23	2010 年 8 月	3539.7	（2970.882，3132.478）	2970.882
24	2010 年 9 月	3616.5	（2870.238，3200.442）	3035.34

参考文献：

[1]牛东晓.电力负荷预测技术及其应用[M].第一版.北京：中国电力出版社，1998.

[2]刘克.实用马尔可夫决策过程[M].北京：清华大学出版社，2004.

[3]唐娜，桂预风，李宝.灰色马尔可夫模型应用于股指分析[C]//第五届中国不确定系统年会论文集，2007：195-198.

[4]王会青，王婷，谷志红.灰色马尔可夫链在高峰负荷预测中的应用[J].电力需求侧管理，2004，6（6）：13-15.