

House Price Forecast Based on Grey Fractional Markov Model

Yixuan Yang¹, Lang Yu¹, Hai Li²

¹School of Science, Southwest University of Science and Technology, Mianyang Sichuan

²School of Civil Engineering and Architecture, Southwest University of Science and Technology, Mianyang Sichuan

Email: 1172875043@qq.com, swust2018yl@163.com, 614618609@qq.com

Received: Apr. 10th, 2019; accepted: Apr. 19th, 2019; published: Apr. 26th, 2019

Abstract

According to the nonlinearity and complexity of the house price data, this paper combines the fractional-order cumulative generation operator and the Markov correction model to establish the MKFGM(1,1) model, and forecasts and analyzes the house prices of Chengdu from April 2017 to May 1818. Firstly, the particle swarm optimization algorithm is used to search for the optimal order of the cumulative generation operator in the real number domain; combined with the GM(1,1) model, the FGM(1,1) model after changing the order can be obtained; and a preliminary prediction of Chengdu city's house price is made. Secondly, the relative error between the predicted value and the actual value is divided depending on the states, then the Markov model is used for error correction, and finally the MKFGM(1,1) model is established. Comparing and analyzing the fitting and prediction results of GM(1,1), FGM(1,1) and MKFGM(1,1) models, it can be seen that the MKFGM(1,1) model has a higher accuracy in housing price forecasting.

Keywords

House Price Forecast, Fractional Gray Model, Markov, Prediction Accuracy, PSO

基于灰色分数阶马尔科夫模型的房价预测

杨羿轩¹, 于浪¹, 李海²

¹西南科技大学理学院, 四川 绵阳

²西南科技大学土木工程与建筑学院, 四川 绵阳

Email: 1172875043@qq.com, swust2018yl@163.com, 614618609@qq.com

收稿日期: 2019年4月10日; 录用日期: 2019年4月19日; 发布日期: 2019年4月26日

摘要

针对房价数据的非线性性和复杂性的特点, 本文结合分数阶累加生成算子和马尔科夫修正模型建立了

MKFGM(1,1)模型,并对成都市2017年4月~2018年5月房价进行预测分析。首先,利用粒子群算法在实数域内搜索累加生成算子的最优阶数 r ,结合GM(1,1)模型可以得到改变阶数后的FGM(1,1)模型,并对成都市房价数据做出初步预测。其次,将预测值与实际值的相对误差进行状态划分,然后使用马尔科夫模型进行误差修正,最终建立MKFGM(1,1)模型。对比分析GM(1,1)、FGM(1,1)和MKFGM(1,1)模型的拟合及预测结果,可以看出MKFGM(1,1)模型在房价预测方面有更高的精度。

关键词

房价预测, 分数阶灰色模型, 马尔科夫, 预测精度, PSO

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

在许多国家和地区房地产业已成为国民经济的支柱产业,是经济发展的重要力量。随着我国市场经济的发展,房地产在国民经济中的地位不断提高,影响不断扩大。上世纪90年代中后期,政府明确提出把房地产业作为国民经济新的增长点。房价的高低、升降不仅对经济的发展有很大的影响,同时也影响着人们的生活水平。因而,如何预测房地产的走势,便成为了我们研究的重要问题。

中国对于房价研究起步较晚,主要基于灰色理论和神经网络模型。Zhang等运用GM(1,1)灰色预测方法,建立了沈阳房地产市场的需求和价格预测模型,具有较好的预测精度[1]。Sun等从房价的诸多影响因素中选取主要因子,结合GM(1,N)模型得出了昆明市房价的预测模型[2]。王钧甲全面考虑房价的影响因素,建立了基于混沌理论的BP神经网络模型[3]。王聪对比多元线性回归模型,建立了多因素LOGISTIC回归模型,提高了预测精度[4]。Tang等根据北京市二手房年房价指数非线性变化的特点,将BAT算法引入传统SVR模型,对模型的三个参数进行优化。结合Web搜索数据,建立了BA-SVRWSD混合模型[5]。郑永坤,刘春等利用差分自回归移动平均来排除其他对房价有影响的复杂因素,建立了ARIMA预测模型[6]。前人在房价预测上已经做了很多工作,提出了多种模型,但都存在着一些局限性。灰色GM(1,1)模型所需的样本数据少,但要求数据具有指数增长规律。BP神经网络模型通过输入与输出的非线性映射,对数据具有较高的处理能力,但其训练速率较低,且容易陷入局部最优解。回归预测模型预测过程简单,但预测精度不高,外推特性差。时间序列模型[7]原理简单,但难以反映数据内在因素的影响,容易存在较大的误差。

灰色预测模型作为主要的预测模型之一,还具有很多的研究价值。Wu等于2013年提出将分数阶累加生成算子与灰色预测模型相结合的理论,基于新信息优先原理,借助分数阶蕴含的“in between”思想,通过选择合适的累加阶数使得增高序列变平缓,从而提高预测精度[8]。Ma等使用grey wolf算法从实数范围内搜索最优累加生成算子,扩大了其取值范围,进一步完善了分数阶理论体系[9]。

本文将分数阶累加生成算子与GM(1,1)模型结合,对成都市房价进行短期预测,并借助马尔科夫模型对预测结果进行误差修正,最终建立灰色分数阶马尔科夫预测模型,以下简称MKFGM(1,1)模型。利用2017年4月~2018年3月成都市房价作为建模数据,2018年4、5月房价为预测检验数据。对比分析GM(1,1)模型、FGM(1,1)模型以及本文建立的MKFGM(1,1)模型。结果表明MKFGM(1,1)模型具有更好的拟合效果和更高的预测精度。

2. 灰色分数阶马尔科夫模型

2.1. 分数阶 FGM(1,1)模型

定义 1: 设原始非负序列为 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))$, $r \in R$, 则 $X^{(r)} = (x^{(r)}(1), x^{(r)}(2), x^{(r)}(3), \dots, x^{(r)}(n))$ 是 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))$ 的 r 阶累加生成序列, 其中:

$$x^{(r)}(k) = \sum_{i=1}^k \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)\Gamma(r)} x^{(0)}(i), k=1,2,\dots,n \quad (1)$$

这里 $\Gamma(r+k-i) = (r+k-i-1)!$, $\Gamma(k-i+1) = (k-i)!$, $\Gamma(r) = (r-1)!$ 。

定义 2: 设 $X^{(0)}$ 与 $X^{(r)}$ 如定义 1 所示, $X^{(r)}$ 是 $X^{(0)}$ 的分数阶累加生成序列, $Z^{(r)} = (z^{(r)}(2), z^{(r)}(3), \dots, z^{(r)}(n))$, 其中:

$$z^{(r)}(k) = \frac{x^{(r)}(k) + x^{(r)}(k-1)}{2}, k=2,3,\dots,n \quad (2)$$

称灰色微分方程 $x^{(r)}(k) - x^{(r)}(k-1) + az^{(r)}(k) = b$ 为 r 阶累加灰色 GM(1,1)模型, 即为 FGM(1,1)模型。特别的, 当阶数 $r=1$ 时为 GM(1,1)模型。

FGM(1,1)模型 $x^{(r)}(k) - x^{(r)}(k-1) + az^{(r)}(k) = b$ 中 $\hat{u} = [a, b]^T$ 可以运用最小二乘法估计:

$$\hat{u} = (B^T B)^{-1} B^T Y \quad (3)$$

其中 Y, B 分别为:

$$Y = \begin{bmatrix} x^{(r-1)}(2) \\ x^{(r-1)}(3) \\ \vdots \\ x^{(r-1)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(r)}(2) & 1 \\ -z^{(r)}(3) & 1 \\ \vdots & \vdots \\ -z^{(r)}(n) & 1 \end{bmatrix} \quad (4)$$

其中, $x^{(r-1)}(k) = x^{(r)}(k) - x^{(r)}(k-1), k=2,3,\dots,n$ 。

定义 3: 参数如定义 2 所述, 则称

$$\frac{dx^{(r)}(t)}{dt} + ax^{(r)}(t) = b \quad (5)$$

为 r 阶累加灰色 FGM(1,1)模型 $x^{(r)}(k) - x^{(r)}(k-1) + az^{(r)}(k) = b$ 的白化微分方程。设 B, Y, \hat{u} 如定义 2 所述, 则求解微分方程(3), 可以得到模型 $x^{(r)}(k) - x^{(r)}(k-1) + az^{(r)}(k) = b$, 令 $\hat{x}^{(0)}(1) = x^{(0)}(1)$, 得到时间序列响应式为:

$$\hat{x}^{(r)}(t) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-a(t-1)} + \frac{b}{a}, t=1,2,\dots,n \quad (6)$$

离散后的得到:

$$\hat{x}^{(r)}(k) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-a(k-1)} + \frac{b}{a}, k=1,2,\dots,n \quad (7)$$

还原后的值为:

$$\hat{x}^{(0)}(k) = \left(\hat{x}^{(r)}\right)^{(-r)}(k) = \sum_{i=0}^{k-1} (-1)^i \frac{\Gamma(r+1)}{\Gamma(i+1)\Gamma(r-i+1)} \hat{x}^{(r)}(k-i), k=2,3,\dots,n \quad (8)$$

其中 $\Gamma(r+1) = r!, \Gamma(i+1) = i!, \Gamma(r-i+1) = (r-i)!$ 。

2.2. 马尔科夫模型

马尔科夫过程是一类重要的随机过程。它通过对预测对象不同状态的初始概率和状态之间的转移概率的研究, 确定预测对象未来状态的变化趋势。马尔科夫过程适合处理波动的数据且具有无后效性和短期预测效果良好的特点, 已被广泛运用到军事、生物、气象等领域[10] [11] [12]。

本文运用马尔科夫模型对 FGM(1,1)模型的预测值与实际值的相对误差进行状态划分。依据数据序列的状态可以构造状态转移矩阵, 从而能够预测出后一个数据最可能所处的误差状态。根据误差状态划分的上下限, 即可得到马尔科夫模型的修正值。马尔科夫模型可表示为:

$$C_{h+1} = C_h * P \quad (9)$$

其中, C_{h+1}, C_h 分别为在 $h+1$ 、 h 时刻的状态概率向量, P 为一步状态转移概率矩阵。

2.2.1. 划分预测状态

根据原始数据与 FGM(1,1)模型预测数据的相对误差, 可以将误差划分为若干个状态。状态的数量没有严格的限制, 一般由原始数据的多少和数据误差的范围大小来确定。状态区间为:

$$E_i = [Q_{i1}, Q_{i2}], [i=1, 2, \dots, u] \quad (10)$$

其中, Q_{i1}, Q_{i2} 分别表示状态区间相对误差的上、下限, u 为状态划分数量。

2.2.2. 构造状态转移概率矩阵

由状态 E_i 经过一个时期转移到 E_j 的一步转移概率为 P_{ij} , 将所有状态的一步转移概率构成一个矩阵, 即:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1u} \\ P_{21} & P_{22} & \cdots & P_{2u} \\ \vdots & \vdots & & \vdots \\ P_{u1} & P_{u2} & \cdots & P_{uu} \end{bmatrix} \quad (11)$$

其 v 步转移概率矩阵为 $P^{(v)} = P^{(v-1)}P = P^v$, $v=1, 2, \dots, u$ 。

2.2.3. 确定预测值

选取离预测数据最近的 u 组数据, 按照由近到远的顺序确定步数 t 为 $1, 2, 3, \dots, u$, 然后取各数据对应的 t 步状态转移矩阵的行向量构成新的矩阵, 由新矩阵的列向量之和确定预测值最可能所处的状态。状态确定以后, 状态区间也就确定了, 马尔可夫修正值为该区间的中点, 即 $\frac{Q_{i1} + Q_{i2}}{2}$, 最终得到的预测值为:

$$y(k) = \frac{\hat{x}^{(r)}(k)}{1 + (Q_{i1} + Q_{i2})} \quad (12)$$

2.3. 模型误差检验准则

为了检验本文中模型的预测精度, 本文采用平均绝对百分误差 MAPE 来描述拟合预测数据与真实数据的误差。根据模型得出的数据与实际数据可以计算出用于建模的拟合数据(Fitting data)的平均绝对误差(MAPEF), 建模后预测数据(Predicted data)的平均绝对误差(MAPEP)和总数据(Total data)的平均绝对误差 MAPET, 其表达式分别如下:

$$\text{MAPEF} = \frac{100}{m} \sum_{i=1}^m \left| \frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \quad (13)$$

$$\text{MAPEP} = \frac{100}{n-m} \sum_{i=m+1}^n \left| \frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \quad (14)$$

$$\text{MAPET} = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \quad (15)$$

其中, m 是用于预测模型的样本个数, n 是样本数据的总个数。

2.4. 最优阶数的确定

使用分数阶 FGM(1,1)模型对原始数据进行预测时, 首先要确定累加生成算子的阶数 r , 然后利用式(2)求出参数 (a, b) 。最后利用时间响应方程求出原始数据预测值 $\hat{x}^{(0)}(k)$ 。为了找到最优阶数 r , 本文采用 MAPET 作为目标函数, 形式如下:

$$\begin{aligned} \min \text{MAPET} &= \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{x}^{(0)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \right| \\ \text{s.t.} &\begin{cases} r \in R, \\ k = 2, 3, \dots, n \end{cases} \end{aligned} \quad (16)$$

由于以上公式具有非线性性, 直接利用公式寻找最优阶数十分困难。为此, 本文使用粒子群算法对分数阶算子进行计算。

粒子群算法(PSO)是一种模拟自然界中鸟集群飞行觅食的行为方式演化而来的随机搜索方法。它于1995年由 James Kennedy 和 Russell Eberhart 共同提出[13]。两人于2001年出版《群体智能》将粒子群算法的影响扩大, 引起了国内外学者的广泛关注[14]。粒子群算法模拟鸟类的觅食行为, 将求解问题的搜索空间类比成鸟类的飞行空间, 将每一只鸟抽象成一个没有质量和体积的粒子, 用来表征问题的一个可行解, 最后将寻找最优解的过程看作鸟类寻找食物的过程, 进而求解复杂的优化问题。

粒子群算法作为一种新兴的智能优化算法, 主要优点是有很好的并行性和全局寻优能力, 并且不存在求导和函数连续性的限定。粒子群算法参数少而容易实现, 对于非线性问题和多峰问题都有很好的全局搜索能力。近年来, 粒子群算法已被广泛应用于机器学习、函数优化、能源预测、模式分类、自适应控制等领域。本文粒子群算法以 MAPET 最小为目标, 寻找最优阶数, 其算法步骤如下:

- 1) 初始化群体粒子个数, 粒子维数等。
- 2) 初始化种群粒子位置和速度, 计算每个粒子的适应度值, 分别与个体极值和全局极值的最优位置及最优值比较, 择优选取。
- 3) 迭代更新粒子的速度和位置, 进行边界条件处理。
- 4) 判断是否满足终止条件; 若不满足, 则继续搜索; 若满足, 则将该粒子作为粒子群算法的最优计算结果。

3. 成都市房价预测分析

3.1. 样本数据

本文实验所需的数据是通过 python 爬虫从中国各大房产网站爬取的。这些房产网站都是开放的平台,

上面集合了各个地区的房源信息, 过滤掉虚假和虚高的房价, 最后得出的均价。经过笔者的对比观察, 各房产网站的价格几乎一致, 具有较高的可信度。最终, 本文使用安居客房产网站上成都市 2017 年 4 月至 2018 年 5 月历史均价数据作为实验研究基础。

3.2. FGM(1,1)模型预测

将以上所得数据导入 matlab 程序中, 利用粒子群算法得出累加生成算子的最优阶数为-0.37。则 FGM(1,1)模型为 GM-0.37(1,1)。并且, 由最小二乘法得出参数 $a = -0.1513$, $b = 27912.0302$, 时间响应式为:

$$\hat{x}^{(-0.37)}(k) = 193884.364e^{0.1513(k-1)} - 184481.364, k = 1, 2, \dots, 12 \quad (17)$$

根据累减还原式(4), 当 $k = 1, 2, \dots, 12$ 时, 则可计算得到预测值。另外, 可取阶数 r 为 1, 得到 GM(1,1)模型的预测值。将 GM(1,1)模型与 FGM(1,1)模型的预测值和实际值相比, 结果如表 1 所示:

Table 1. Comparison data between GM(1,1) and FGM(1,1) models.

表 1. GM(1,1)和 FGM(1,1)模型之间的比较数据

时间	实际值/元	GM(1,1)模型/元	相对误差/%	FGM(1,1)模型/元	相对误差/%
2017.4	9403	9403.00	0.000	9403.00	0.000
2017.5	9636	9475.26	-1.668	9844.96	2.169
2017.6	9839	9779.23	-0.608	9895.98	0.579
2017.7	9992	10,092.94	1.010	10,081.00	0.891
2017.8	10,185	10,416.72	2.275	10,383.59	1.950
2017.9	10,562	10,750.89	1.788	10,746.87	1.750
2017.10	11,068	11,095.77	0.251	11,132.79	0.585
2017.11	11,617	11,451.72	-1.423	11,521.00	-0.826
2017.12	12,034	11,819.09	-1.786	11,901.70	-1.099
2018.1	12,375	12,198.25	-1.428	12,270.69	-0.843
2018.2	12,542	12,589.56	0.379	12,626.58	0.674
2018.3	12,825	12,993.43	1.313	12,969.27	1.125
2018.4	13,036	13,410.26	2.871	13,299.26	2.020
2018.5	13,147	13,840.46	5.275	13,617.32	3.577

3.3. 马尔科夫模型修正

3.3.1. 对每个月进行状态区间划分

根据 FGM(1,1)模型的相对误差划分状态区间, 从表 1 可以看出 FGM(1,1)模型的前 12 个拟合数据相对误差的最小值为-1.099%, 最大值为 2.169%。由此, 根据等间距规则划分四个状态区间, 分别为 $E_1(-1.10, -0.275]$, $E_2(-0.275, 0.55]$, $E_3(0.55, 1.375]$, $E_4(1.375, 2.20]$, 如表 2 所示:

Table 2. State partition

表 2. 状态划分区间

状态	E_1	E_2	E_3	E_4
相对误差	$(-1.10, -0.275]$	$(-0.275, 0.55]$	$(0.55, 1.375]$	$(1.375, 2.20]$

3.3.2. 构建状态转移矩阵

根据每个数据所处的状态，以及由当前状态到下一个状态的概率，得到如下的 1、2、3、4 步状态转移概率矩阵：

$$P = \begin{bmatrix} 0.6667 & 0 & 0.3333 & 0 \\ 0 & 0 & 0 & 1.0000 \\ 0.2500 & 0 & 0.5000 & 0.2500 \\ 0 & 0 & 0.6667 & 0.3333 \end{bmatrix}, P^{(2)} = \begin{bmatrix} 0.5278 & 0 & 0.3889 & 0.0833 \\ 0 & 0 & 0.6667 & 0.3333 \\ 0.2917 & 0 & 0.5000 & 0.2083 \\ 0.1667 & 0 & 0.5556 & 0.2778 \end{bmatrix}$$

$$P^{(3)} = \begin{bmatrix} 0.4491 & 0 & 0.4259 & 0.1250 \\ 0.1667 & 0 & 0.5556 & 0.2778 \\ 0.3194 & 0 & 0.4861 & 0.1944 \\ 0.2500 & 0 & 0.5185 & 0.2315 \end{bmatrix}, P^{(4)} = \begin{bmatrix} 0.4059 & 0 & 0.4460 & 0.1481 \\ 0.2500 & 0 & 0.5185 & 0.2315 \\ 0.3345 & 0 & 0.4792 & 0.1863 \\ 0.2963 & 0 & 0.4969 & 0.2068 \end{bmatrix}$$

3.3.3. 计算预测值

选择最近的几组数据构建新的状态转移矩阵，见表 3：

Table 3. State transition matrix predicting the state of April 2018

表 3. 预测 2018 年 4 月所属状态的状态转移矩阵

月份	起始状态	转移步数	E_1	E_2	E_3	E_4
2018.3	E_3	1	0.25	0	0.5	0.25
2018.2	E_3	2	0.2917	0	0.5	0.2083
2018.1	E_1	3	0.4491	0	0.4259	0.1250
2017.12	E_1	4	0.4059	0	0.4460	0.1481
合计			1.3967	0	1.8719	0.7314

通过表 3 可知，2018 年 4 月最有可能所属的状态是 E_3 ，并且本月 FGM(1,1)模型的预测值为 13,172.479，根据公式(12)可得马尔科夫模型预测值为 13,487.499。

在进行 2018 年 5 月数据预测时，可以将 2018 年 4 月数据按已知量计算，按照上述方法，得出 2018 年 5 月预测值为 13,487.499。MKFGM(1,1)模型 2017.4~2018.5 的拟合及预测数据见表 4：

Table 4. Comparison data between FGM(1,1) and MKFGM(1,1) models

表 4. FGM(1,1)和 MKFGM(1,1)模型之间的比较数据

时间	实际值/元	GM(1,1)模型/元	相对误差/%	FGM(1,1)模型/元	相对误差/%
2017.4	9403	9403.00	0.000	9403.000	0.000
2017.5	9636	9844.96	2.169	9672.074	0.374
2017.6	9839	9895.98	0.579	9801.638	-0.380
2017.7	9992	10,081.00	0.891	9984.897	-0.071
2017.8	10,185	10,383.59	1.950	10,284.605	0.978
2017.9	10,562	10,746.87	1.750	10,644.415	0.780
2017.10	11,068	11,132.79	0.585	11,026.663	-0.373
2017.11	11,617	11,521.00	-0.826	11,411.169	-1.772
2017.12	12,034	11,901.70	-1.099	11,984.090	-0.415

Continued

2018.1	12,375	12,270.69	-0.843	12,355.639	-0.156
2018.2	12,542	12,626.58	0.674	12,713.987	1.371
2018.3	12,825	12,969.27	1.125	12,845.626	0.161
2018.4	13,036	13,299.26	2.020	13,172.479	1.047
2018.5	13,147	13,617.32	3.577	13,487.499	2.590

从表 4 可以看出 FGM(1,1)模型已经具有一定的效果, 预测结果良好, 而本文建立的 MKFGM(1,1)模型则更加精确。

为了比较 GM(1,1)模型、FGM(1,1)模型、MKFGM(1,1)模型的预测精度, 以表 4 中前 12 个数据为拟合数据, 后 2 个数据为预测数据, 根据式(6)、(7)、(8)分别计算 3 种模型的 MAPEF、MAPEP 和 MAPET, 结果见表 5。另外, 为了更直观地展示 3 种模型的预测效果, 图 1 (左图)绘制了 GM(1,1), FGM(1,1)和 MKFGM(1,1)模型的拟合预测曲线, 图 1 (右图)给出了 3 种模型的相对误差的绝对值的对比情况。

Table 5. The average absolute percentage error of the fitting, prediction and total data for the three models

表 5. 三种模型的拟合、预测、总数据的平均绝对百分误差

模型误差/%	GM(1,1)	FGM(1,1)	MKFGM(1,1)
MAPEF	1.161	1.041	0.569
MAPEP	4.073	2.798	1.818
MAPET	1.577	1.292	0.748

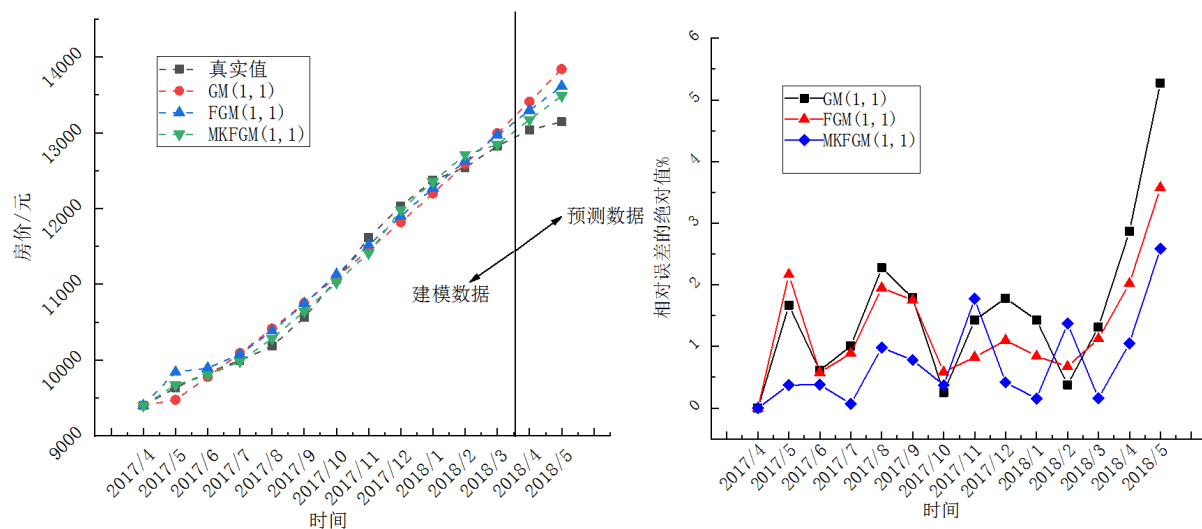


Figure 1. Three models fitting prediction comparison chart

图 1. 三种模型拟合预测对比图

从表 5 可以看出, GM(1,1)模型的 MAPEF、MAPEP、MAPET 分别为 1.161%、4.073%、1.577%, FGM(1,1)模型的 MAPEF、MAPEP、MAPET 较好, 分别为 1.041%、2.798%、1.292%, MKFGM(1,1)模型的 MAPEF、MAPEP、MAPET 分别为 0.569%、1.818%、0.748%。其中, MKFGM(1,1)模型的 MAPEF、MAPEP、MAPET 均最低, 说明 MKFGM(1,1)模型的拟合数据波动最小, 拟合程度最高, 并且预测结果也最好。从图 1 的左图和右图也可以直观地看出 MKFGM(1,1)模型的建模数据和预测数据均最贴合实际数据, 并且其相对误差的绝对值波动范围最小。

4. 结论

灰色预测模型可以通过少量的、信息匮乏的数据进行建模,从而找出数据内部隐含的规律,并对数据进行短期预测。本文使用粒子群算法搜索 FGM(1,1)模型的累加生成算子的最优阶数,初步改进了 GM(1,1)模型。再使用马尔科夫模型对 FGM(1,1)模型的预测结果进行误差修正,得到了本文的 MKFGM(1,1)模型。对比分析 GM(1,1)、FGM(1,1)和 MKFGM(1,1)模型的拟合及预测结果,可以得到以下结论:

1) 三个模型的 MAPEF、MAPEP、MAPET 均在 5%以下,说明这三个模型应用于房价预测都是切实有效的。

2) 本文建立的 MKFGM(1,1)模型的预测精度最高,MAPET 仅为 0.748%,远低于 GM(1,1)模型和 FGM(1,1)模型,改进效果显著。

3) MKFGM(1,1)模型作为一种优秀且稳定的预测模型,可考虑将其进一步应用在金融、能源、交通运输等领域。

基金项目

西南科技大学理学院创新基金项目“基于 scrapy 的房地产数据爬虫系统”(项目编号: LXCX-19),主持人: 杨羿轩。

参考文献

- [1] Zhang, S., Gao, M. and Wang, X. (2010) The Real Estate Market Forecast of Shenyang Based on Gray System Theory. *International Conference on Management & Service Science*, Wuhan, 24-26 August 2010, 1-4.
- [2] Sun, Y.M. and Wu, L.G. (2013) Case Analysis of GM(1,N) Model in Predicting Kunming Real Estate Prices. *Applied Mechanics and Materials*, **444-445**, 1781-1786. <https://doi.org/10.4028/www.scientific.net/AMM.444-445.1781>
- [3] 王钧甲. 基于混沌理论的沈阳普通商品住宅价格研究[D]: [硕士学位论文]. 沈阳: 沈阳建筑大学, 2014.
- [4] 王聪. 基于多因素 LOGISTIC 的城市房地产价格预测模型研究[D]: [硕士学位论文]. 大连: 大连理工大学, 2008.
- [5] Tang, X., et al. (2018) Research on Forecast of Second-Hand House Price in Beijing Based on SVR Model of Bat Algorithm. *Statistical Research*, **35**, 71-81.
- [6] 郑永坤, 刘春. 基于 ARIMA 模型的二手房价格预测[J]. *计算机与现代化*, 2018(4): 122-126.
- [7] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008) *Time Series Analysis: Forecasting and Control*. 4th Edition, Wiley, Hoboken.
- [8] Wu, L., Liu, S., Yao, L., et al. (2013) Grey System Model with the Fractional Order Accumulation. *Communications in Nonlinear Science and Numerical Simulation*, **18**, 1775-1785. <https://doi.org/10.1016/j.cnsns.2012.11.017>
- [9] Ma, X., Xie, M., Wu, W., et al. (2019) The Novel Fractional Discrete Multivariate Grey System Model and Its Applications. *Applied Mathematical Modeling*, **70**, 402-424. <https://doi.org/10.1016/j.apm.2019.01.039>
- [10] Şahingil, M.C. and Yurttas, R. (2012) The Determination of Flare Launching Programs to Use against Pulse Width Modulating Guided Missile Seekers via Hidden Markov Models.
- [11] Krogh, A., Larsson, B., Von, G.H., et al. (2001) Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *Journal of Molecular Biology*, **305**, 567-580. <https://doi.org/10.1006/jmbi.2000.4315>
- [12] Thyer, M. and Kuczera, G. (2000) Modeling Long-Term Persistence in Hydroclimatic Time Series Using a Hidden State Markov Model. *Water Resources Research*, **36**, 3301-3310. <https://doi.org/10.1029/2000WR900157>
- [13] Kennedy, J. and Eberhart, R. (1995) Particle Swarm Optimization. *IEEE International Conference on Neural Networks*, Perth, Vol. 4, 1942-1948.
- [14] Kennedy, J. and Eberhart, R.C. (2001) *Swarm Intelligence*. Academic Press, Cambridge.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：csa@hanspub.org