

Feasibility Study on the Forecast of China's Small and Medium Board Stock Index

—Based on GM(1,1) and GM(2,1) Gray Models

Lingsen Kong, Huqin Yan

Xiamen National Accounting Institute, Xiamen Fujian
Email: 498312408@qq.com

Received: May 15th, 2020; accepted: May 28th, 2020; published: Jun. 4th, 2020

Abstract

This article uses the GM(1,1) and GM(2,1) gray prediction models and selects the transaction data of the SME Composite Index (399101) from February 6, 2020 to April 30, 2020 to compare the prediction effects of the two models horizontally and vertically. The prediction results show that the gray system model is more suitable for short-term and more monotonic data samples for the prediction of China's stock index, which can better fit the law of stock index changes. For long-term or non-monotonic data, the gray system model has a poor fitting effect on the stock index and can't provide a reference for the stock market price prediction.

Keywords

Gray Prediction Model, Stock Forecast, SME Composite Index

对我国中小板股指预测的可行性研究

——基于GM(1,1)和GM(2,1)灰色模型

孔令森, 阎虎勤

厦门国家会计学院, 福建 厦门
Email: 498312408@qq.com

收稿日期: 2020年5月15日; 录用日期: 2020年5月28日; 发布日期: 2020年6月4日

摘要

本文采用GM(1,1)和GM(2,1)灰色预测模型, 随机选取了2020年2月6日至2020年4月30日的中小板综合

指数(399101)交易数据, 横向和纵向比较两种模型的预测效果, 实证分析证明灰色模型预测股指的可行性。预测结果表明: 灰色系统模型对于我国股指的预测更多地适用于短期且摆动变化较为单调的数据样本, 这样才能较好地拟合出股指变动的规律, 就长期或者数据摆动变化非单调而言, 灰色系统模型对于股指的拟合效果较差, 不能为股票市场的价格预测提供参考。

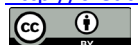
关键词

灰色预测模型, 股票预测, 中小板综合指数

Copyright © 2020 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

改革开放以来, 我国经济快速增长, 一举成为第二大经济体, 在中国经济增长过程中, 股票市场对于经济增长做出了卓越的贡献, 主要表现在我国证券行业和上市公司对于 GDP 的贡献。股票市场的稳定关乎金融市场稳定, 更关乎我国宏观经济的稳定。以深交所为例, 深证成指、中小板指数、创业板指数等是评价相应版块的重要信息, 因此, 其变动备受投资者和股民的关注, 但是股票市场受多种因素影响, 存在着不确定性和复杂性, 具体表现在股票价格的波动频繁和剧烈, 这样一来, 相关指数的有效预测尤为重要。灰色预测模型是灰色系统理论的重要组成部分, 我们可以将整个股指系统看作为灰色系统, 利用大数据和 python 进行数据处理, 并通过 GM(1,1)和 GM(2,1)灰色预测模型对中小板综合指数的变动趋势进行挖掘, 去验证模型的可行性和有效性。

2. 文献综述

从春霞[1]在提出灰色预测模型是以 GM(1,1)模型为依托, 对股价进行数列和灾变预测, 其中数列预测根据不同的时期, 比如日、月、年, 进行股指数值的预测, 而灾变预测则是对股指可能发生的异常变动进行预测。张宇敬[2]运用灰色 GM(1,1)模型对三只上市公司的股价进行为期 3 天的预测, 研究表明, 该模型仅局限于短期的预测, 不适合长期预测, 在股价受到不确定性因素或系统性影响时, 模型预测的结果会极大地偏离实际。赵春阁[3]等人将灰色模型和 MATLAB 相结合对企业利润进行预测, 证明了灰色模型适用于大数据的预测, 并且在数据较少时测试结果依然准确。赵新冀[4]等人研究发现 GM(2,1)模型对非单调变化的原始数据序列进行预测时, 多数情况下效果并不理想, 针对此现状提出改进方法。Ning Xu [5]等人在对 GM(2,1)模型的预测性研究中得出改进后的 GM(2,1)模型更适用于短期的工程项目预测。

3. 灰色系统模型

灰色系统理论是中国学者邓聚龙教授 1982 年 3 月在国际上首先提出来的, 灰色系统理论认为在大量繁琐的数据背后, 一定蕴含着某种潜在的规律, 生成的灰数可以帮我们获取信息, 然后系统通过获取的信息去寻找内部的变动规律, 从而生成相对应的样本序列集, 再将其代入 python 中设定的模型去对样本的发展趋势进行预测, 并检验预测值和样本的拟合效果。常见的灰色模型有 GM(1,1)邓聚龙估计、GM(1,1)微分方程直接求解、GM(2,1)白化微分方程等, 在下面, 我将选用 GM(1,1)微分方程直接求解和 GM(2,1)白化微分方程两种模型分别进行研究和对比。

1) GM(1,1)模型

GM(1,1)模型是最简单的灰色系统模型, 其基础是常微分方程的解析解, GM(1,1)微分方程直接求解的建立过程如下:

假设有一个变量 $x(t)$ 是时间变量 t 的函数, 它满足一阶常微分方程条件:

$$\frac{dx(t)}{dt} + ax(t) = b \quad (1)$$

这里, 参数 a 和 b 是两个常系数。假设参数 C 是任意常数, 那么, 该微分方程的解析解或者通解为:

$$x(t) = \frac{b + e^{a(C-t)}}{a} \quad (2)$$

如果 $x(t)$ 有一个初值, 在 $t=0$ 时, 初值为 $x(0)$, 那么, $x(0)$ 也满足这个解, 代入微分方程的通解, 就有:

$$x(t) = \frac{b}{a} + \left(x(0) - \frac{b}{a} \right) e^{-at} \quad (3)$$

该函数具有指数函数的特征, 由于其对于参数 a 和 b 的依赖性很强, 而参数 a 和 b 又是未知的, 且在正常情况下不易估计, 因此, 该模型也被称为灰色系统。由于只有一个变量 $x(t)$, 且是一阶微分, 所以记为 GM(1,1)。

2) GM(2,1)模型

GM(2,1)和 GM(1,1)的区别在于 GM(2,1)是二阶常微分方程, GM(1,1)为一阶常微分方程, GM(2,1)白化微分方程模型建立过程如下:

假设变量 $x(t)$ 是关于时间 t 的二阶常微分方程的解, 满足条件:

$$\frac{d^2x(t)}{dt^2} + a \frac{dx(t)}{dt} + bx(t) = 0 \quad (13)$$

则该微分方程的通解为:

$$x(t) = C_1 e^{\frac{(-a - \sqrt{a^2 - 4b})t}{2}} + C_2 e^{\frac{(-a + \sqrt{a^2 - 4b})t}{2}} \quad (14)$$

为了编程方便, 我们对该关系式予以简化:

$$x(t) = C_1 e^{mt} + C_2 e^{nt} \quad (15)$$

这里, 存在关系式:

$$m = \frac{-a - \sqrt{a^2 - 4b}}{2} \quad (16)$$

$$n = \frac{-a + \sqrt{a^2 - 4b}}{2} \quad (17)$$

$$C_1 = \frac{x(0)e^n - x(1)}{e^n - e^m} \quad (18)$$

$$C_2 = \frac{x(1) - x(0)e^m}{e^n - e^m} \quad (19)$$

4. 实证分析

1) 数据的选取

本文选取中小板综合指数(399101) 2020年2月6日至2020年4月30日共计60日收盘指数作为变量研究对象(表1), 并将全部样本分为三个时间区间, 分别为5日、9日及30日, 将三个区间的数据分别代入Python系统SymPy模块中的GM(1,1)模型和GM(2,1)模型进行横向和纵向的比较分析。同时, 我们对选取的数据作出几点说明: ① 样本数量可以满足测试的需要; ② 样本数据来源真实; ③ 选取的样本期间市场未发生巨大动荡; ④ 样本数据具备时效性。

Table 1. Sample data

表 1. 样本数据

Number	中小综合	Number	中小综合	Number	中小综合
1	9863.877	21	10,942.230	41	9621.553
2	9893.113	22	10,882.445	42	9580.405
3	9993.520	23	10,460.722	43	9893.416
4	9994.733	24	10,700.779	44	9874.985
5	10,150.053	25	10,536.378	45	9949.142
6	10,067.529	26	10,297.588	46	9729.233
7	10,100.085	27	10,192.391	47	9640.415
8	10,421.843	28	9695.416	48	9876.488
9	10,530.491	29	9659.982	49	9836.143
10	10,481.518	30	9533.439	50	9880.677
11	10,704.748	31	9601.004	51	9913.418
12	10,849.843	32	9702.463	52	10,024.014
13	11,020.472	33	9276.544	53	9942.531
14	11,115.367	34	9436.449	54	10,050.901
15	10,790.754	35	9713.964	55	10,003.164
16	10,812.264	36	9612.713	56	9842.426
17	10,286.545	37	9553.202	57	9849.987
18	10,663.261	38	9362.606	58	9819.230
19	10,751.844	39	9430.273	59	9807.414
20	10,765.910	40	9399.161	60	10,004.897

数据来源: 巨潮资讯网。

(数据说明: 编号1的数据代表时间为2020年2月6日, 编号60的数据代表时间为2020年4月30日)。

2) GM(1,1)模型求解

我们以Python为媒介建立GM(1,1)模型分别以随机选取的中小板综合指数2020年2月6日~2020年2月12日(编号1~5)、2020年2月24日~2020年2月28日(编号13~17)以及2020年3月13日~2020年3月19日(编号27~31)三个5日区间数据, 2020年2月6日~2020年2月18日(编号1~9)和2020年4月13日~2020年4月23日(编号47~55)两个9日区间数据, 2020年2月6日~2020年3月18日(编号1~30)和2020年2月20日~2020年4月01日(编号11~40)两个30日区间数据进行回归分析和预测, 得到的相关数据以及预测拟合程度图如表2、图1~7所示:

Table 2. Data fitting effect of GM(1,1)model
表 2. GM(1,1)模型数据拟合效果

GM(1,1)模型数据拟合效果					
时间长度	R^2	R	Adj R^2	SER	RMSE
5日(1~5)	0.871	0.933	0.828	115.435	89.415
5日(13~17)	0.874	0.935	0.932	110.960	85.949
5日(27~31)	0.982	0.991	0.976	45.024	34.876
9日(1~9)	0.830	0.911	0.806	151.046	133.210
9日(47~55)	0.983	0.992	0.981	54.74	48.28
30日(1~30)	0.254	0.504	0.228	361.762	349.495
30日(11~40)	0.581	0.762	0.565	304.514	294.189

其中, 主要的两个指标:

RMSE (Root mean squared error)表示均方根、标准差, 是 MSE 的平方根, 用来评估模型的误差, 开平方可以保持误差的单位级和我们的数据单位级保持一致, R-Squared 的结果用来判断模型的好坏, 其取值范围为[0,1], 如果结果是 0, 说明模型拟合效果很差; 如果结果是 1, 说明模型无错误。一般来说, R-Squared 越大, 表示模型拟合效果越好。R-Squared 反映的是大概有多准, 因为, 随着样本数量的增加, R-Square 必然增加, 无法真正定量说明准确程度, 只能大概定量。

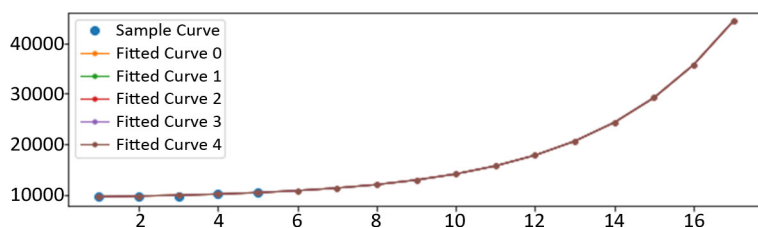


Figure 1. GM(1,1) model 5-day (No.1~5) data prediction diagram

图 1. GM(1,1)模型 5 日(编号 1~5)数据预测图

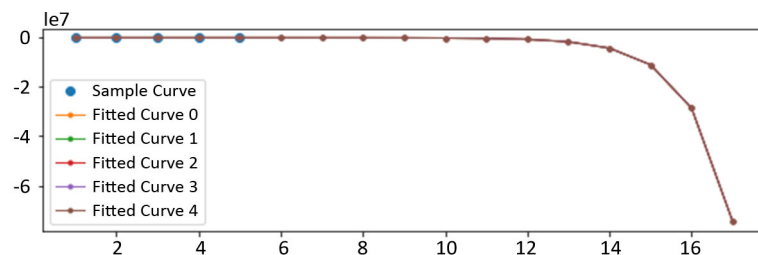


Figure 2. GM(1,1) model 5-day (No.13~17) data prediction diagram

图 2. GM(1,1)模型 5 日(编号 13~17)数据预测图

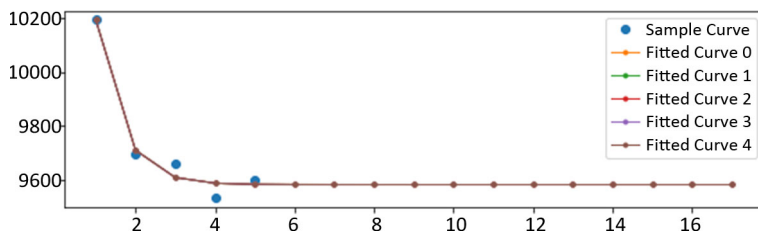


Figure 3. GM(1,1) model 5-day (No.27~31) data prediction diagram

图 3. GM(1,1)模型 5 日(编号 27~31)数据预测图

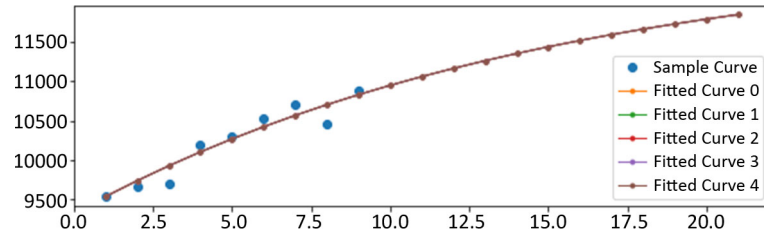


Figure 4. GM(1,1) model 9-day (No.1~9) data prediction diagram
 图 4. GM(1,1)模型 9 日(编号 1~9)数据预测图

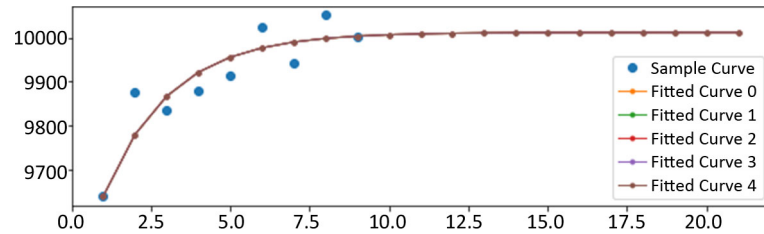


Figure 5. GM(1,1) model 9-day (No.47~55) data prediction diagram
 图 5. GM(1,1)模型 9 日(编号 47~55)数据预测图

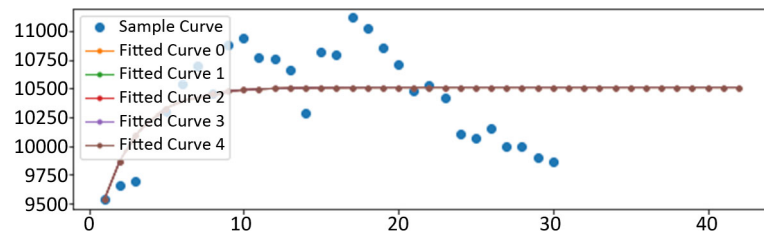


Figure 6. GM(1,1) model 30-day (No.1~30) data prediction diagram
 图 6. GM(1,1)模型 30 日(编号 1~30)数据预测图

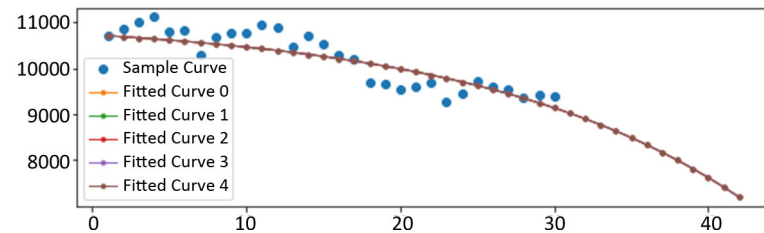


Figure 7. GM(1,1) model 30-day (No.11~40) data prediction diagram
 图 7. GM(1,1)模型 30 日(编号 11~40)数据预测图

从表 2 中可以看出, 时间区间为 5 日时, GM(1,1)模型对于中小板综合指数的三个 5 日样本区间的拟合程度 R^2 达到 0.871、0.874 以及 0.982, 这些都是较好的拟合效果, 同时, 两个 9 日的样本区间也表现出的拟合效果良好, 但是, 随着选取的数据时间区间的进一步扩大, 拟合程度 R^2 出现下降, 在时间区间为 30 日时, 相较于 5 日和 9 日的时间区间出现较大的拟合程度波动, 拟合程度 R^2 下降为 0.581 和 0.254, 有一个区间较短样本表现为断层式的下降, 拟合程度 R^2 为 0.254, 拟合精度极低, 从图 6 中也可以看出样本点并未紧紧的围绕在预测线周围, 模型研究结果告诉我们 GM(1,1)模型对我国中小板指数的短期预测效果要远远高于长期预测, 且长期预测的效果较差, 并不适用于长期预测。

3) GM(2,1)模型求解

我们以 Python 为媒介建立 GM(2,1)模型并分别以随机选取的中小板综合指数 2020 年 2 月 6 日~2020

年2月12日(编号1~5)、2020年2月24日~2020年2月28日(编号13~17)以及2020年3月13日~2020年3月19日(编号27~31)三个5日区间数据,2020年2月6日~2020年2月18日(编号1~9)和2020年4月13日~2020年4月23日(编号47~55)两个9日区间数据,2020年2月6日~2020年3月18日(编号1~30)和2020年2月20日~2020年4月01日(编号11~40)两个30日区间数据进行回归分析和预测,得到的相关数据以及预测拟合程度图如表3、图8~14所示:

Table 3. Data fitting effect of GM(2,1) model

表 3. GM(2,1)模型数据拟合效果

GM(2,1)模型数据拟合效果					
时间长度	R^2	R	Adj R^2	SER	RMSE
5日(1~5)	0.836	0.915	0.782	118.455	91.755
5日(13~17)	0.858	0.926	0.810	130.685	101.229
5日(27~31)	0.970	0.985	0.960	47.896	37.100
9日(1~9)	0.989	0.994	0.988	167.562	339.577
9日(47~55)	0.849	0.921	0.827	44.985	39.673
30日(1~30)	0.930	0.965	0.928	624.718	603.535
30日(11~40)	0.537	0.733	0.520	292.005	282.103

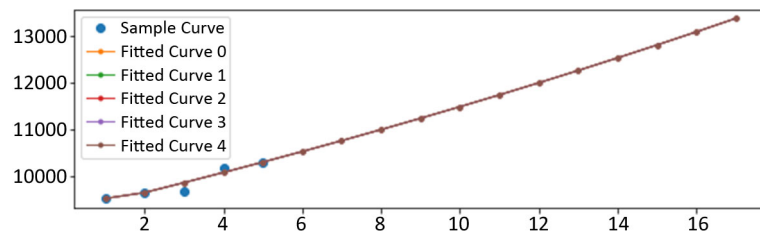


Figure 8. GM(2,1) model 5-day (No.1~5) data prediction diagram

图 8. GM(2,1)模型 5日(编号1~5)数据预测图

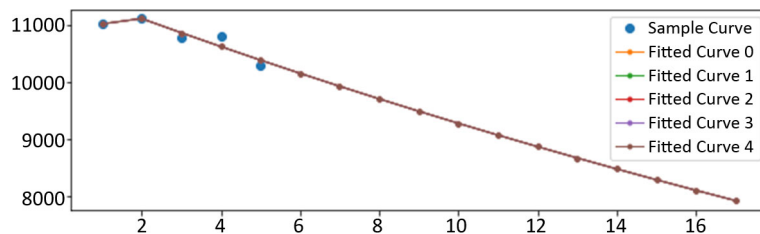


Figure 9. GM(2,1) model 5-day (No.13~17) data prediction diagram

图 9. GM(2,1)模型 5日(编号13~17)数据预测图

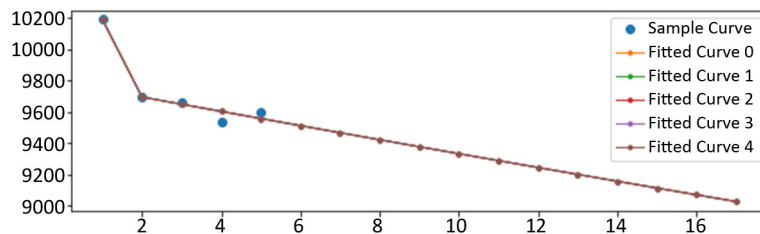


Figure 10. GM(2,1) model 5-day (No.27~31) data prediction diagram

图 10. GM(2,1)模型 5日(编号27~31)数据预测图

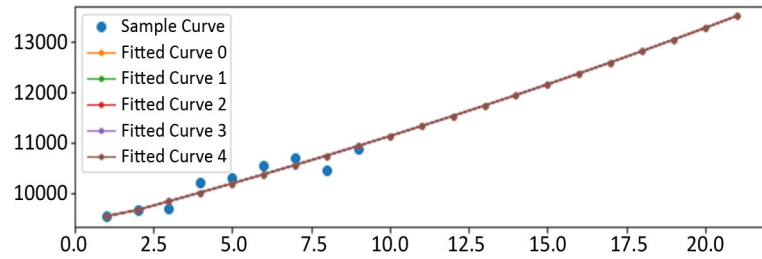


Figure 11. GM(2,1) model 9-day (No.1~9) data prediction diagram

图 11. GM(2,1)模型 9 日(编号 1~9)数据预测图

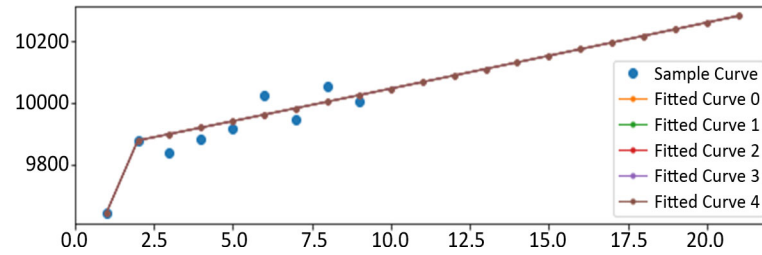


Figure 12. GM(2,1) model 9-day (No.47~55) data prediction diagram

图 12. GM(2,1)模型 9 日(编号 47~55)数据预测图

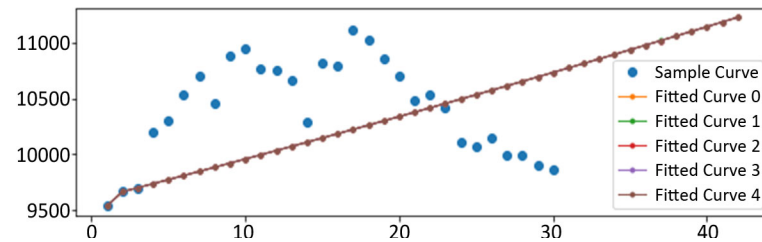


Figure 13. GM(2,1) model 30-day (No.1~30) data prediction diagram

图 13. GM(2,1)模型 30 日(编号 1~30)数据预测图

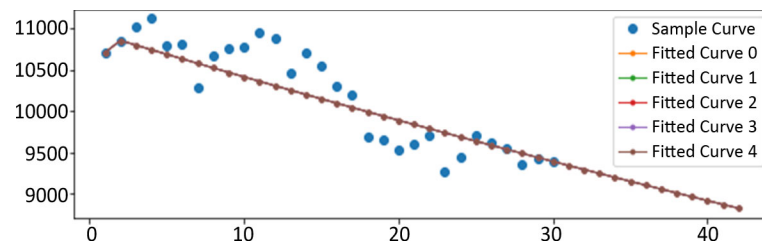
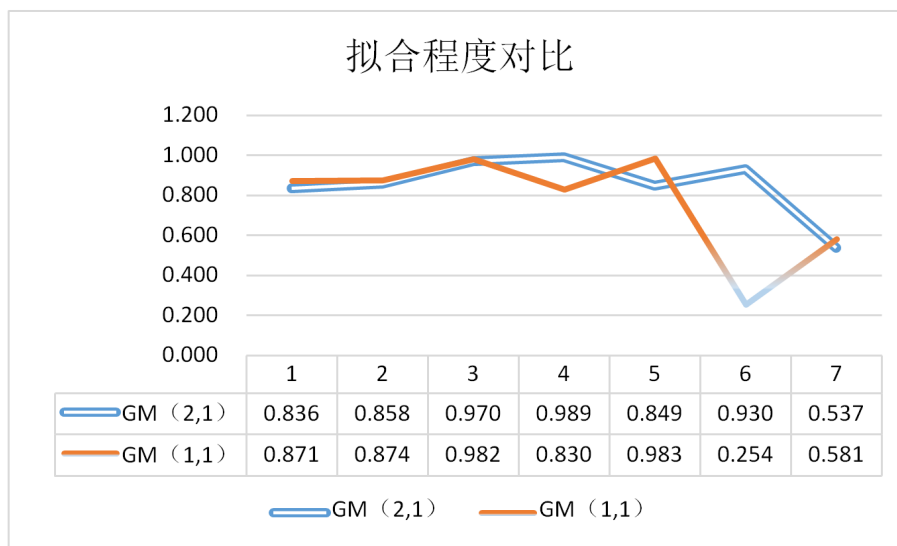


Figure 14. GM(2,1) model 30-day (No.11~40) data prediction diagram

图 14. GM(2,1)模型 30 日(编号 11~40)数据预测图

从表 3 中可以看出, 时间区间为 5 日时, GM(2,1)模型对于中小板综合指数的三个 5 日样本区间的拟合程度 R^2 达到 0.836、0.858 以及 0.970, 样本点均匀的紧紧围绕在预测曲线周围, 这同 GM(1,1)模型对相同数据的拟合预测结果展示一直, 均为较好的拟合效果, 再观察随机选取的两个 9 日区间的拟合效果, 拟合程度 R^2 达到 0.989、0.849, 延续了之前的优良效果, 到此为止, 实证的结果均显示 GM(2,1)模型对于中小板综合指数的短期有很好的预测效果, 可以作为参考, 当选取的数据时间区间变为 30 日时, 拟合程度 R^2 分别为 0.930 和 0.537, 得到的结果表现不一致, 出现相对两极化的情况, 说明数据区间的选取对于模型的预测结果有一定影响, 但总体上看, GM(2,1)模型对于中小板综合指数相同的两个 30 日区间的数据拟合预测效果要优于 GM(1,1)模型, 更具适用性。

5. 结论



数据说明：图 15 中每一列数据为同一时间区间

Figure 15. Comparison diagram of fitting degree

图 15. 拟合程度对比图

通过 GM(1,1)模型和 GM(2,1)模型对不同时间区间的中小板综合指数预测的结果进行横向和纵向的比较可以得出, GM(2,1)模型相比于 GM(1,1)模型具备更优良的纵向预测效果, GM(1,1)模型适用于短期内的预测, 不具备进行长期预测的能力, GM(2,1)模型所适用的时间区间更广泛, 拟合效果更好, 从图 15 中可以看出 GM(2,1)模型的折线大部分在 GM(1,1)模型折线的上方, 并且对于选取的 7 个时间区间 GM(2,1)模型展现出的拟合预测效果更加平稳。另外, 由于股价的变动是随机的、非线性的, 在数据量变大时, 样本非线性变化的特征愈加明显, 这也导致无论是 GM(1,1)模型还是 GM(2,1)模型在对时间区间为 30 日的中小板综合指数进行趋势分析和预测时所得出的对比图拟合性较差。

因此, 通过上述研究可以初步得出这样一个结论, GM(1,1)模型和 GM(2,1)模型在面对非线性、非单调摆动的样本数据时, 均不具备良好的预测效果, 不能为使用者提供有价值的参考。但是, 在样本数目较少时, GM(1,1)模型和 GM(2,1)模型均具备预测能力, 且 GM(2,1)模型的预测拟合效果要更好一些, 可以为投资者对股指的趋势预测提供参考和借鉴。

基金项目

本论文得到了厦门国家会计学院 2019 年“云顶课题: Python 财务数据分析”项目的支持。

参考文献

- [1] 丛春霞, 李秀芳. 灰色预测在股票价格指数预测中的应用[J]. 中国统计, 2000(5): 15-17.
- [2] 张宇敬, 李倩, 蔡虎. 基于灰色预测模型的股价预测研究[J]. 金融教学与研究, 2013(6): 50-52.
- [3] 赵春阁, 徐群. 基于灰色预测 GM(1,1)的企业利润预测[J]. 沈阳大学学报(自然科学版), 2014, 26(5): 382-385.
- [4] 赵新萁, 陈红林. GM(2,1)模型预测公式的改进研究[J]. 武汉理工大学学报, 2006(10): 125-127+131.
- [5] Xu, N., Dang, Y. and Peng, H. (2015) An Optimized Grey GM(2,1) Model and Forecasting of Highway Subgrade Settlement. *Mathematical Problems in Engineering*, 2015, Article ID: 606707. <https://doi.org/10.1155/2015/606707>