

基于 GM(1, 1) 残差尾段修正模型的我国建筑业总产值预测研究

程帅帅, 宋冀龙, 吕海涛

(河北工程大学 管理工程与商学院, 河北 邯郸 056038)

[摘要]在动态灰色预测模型(GM(1, 1)模型)的基础上, 选取我国近七年建筑业总产值作为研究对象, 为了提高模型精度和预测更加准确性, 建立残差尾段修正模型。结果表明, 基于GM(1, 1)残差尾段修正模型预测我国建筑业总产值的准确性有所提高, 且达到相应模型精度要求, 对我国短期建筑业总产值的预测具有一定的参考价值。

[关键词]GM(1, 1); 残差尾段修正; 建筑业总产值

doi:10.3969/j.issn.1673-9477.2018.01.003

[中图分类号] F293.33

[文献标识码] A

[文章编号] 1673-9477(2018)01-007-04

建筑业是指从事房屋、构筑物建造和设备安装活动的法人企业, 是以建筑产品生产为对象的物质生产部门, 是国民经济体系中专门从事土木工程及附属设施的建造, 线路、管道和设备的安装以及装饰装修活动的行业^[1]。建筑业总产值是指在一定时期内完成的以价值表现的生产总量, 是建筑业产品和服务的总和, 反映了建筑业生产成果的综合指标。通过它可以了解建筑业的生产规模、发展速度、经营成果, 并为国家制订经济建设计划提供依据。近年来我国经济发展下行, 呈中高速发展的新常态发展, 建筑业发展面对压力加大、投资增长乏力等复杂局面, 建筑业加快转型升级, 积极推进建筑产业现代化, 攻坚克难, 开拓进取, 整体发展稳中有进^[2]。为了了解我国建筑业总产值动态变化规律, 本文采用动态灰色预测模型(GM(1, 1)模型)预测新常态下我国建筑业总产值的发展变化, 从中探寻一定的规律, 在此模型的基础上对残差尾段进行修正, 模拟出我国建筑业后三年的总产值。

一、动态灰色预测模型

(一) 动态灰色预测模型(GM(1, 1))的建立

设 $x^{(0)} = \{x^{(0)}_{(1)}, x^{(0)}_{(2)}, \Lambda, x^{(0)}_{(n-1)}, x^{(0)}_{(n)}\}$ 为原始离散数据列。

(1) 对 $x^{(0)}$ 作 1-AGO, 即令

$$x^{(1)}_{(k)} = \sum_{i=1}^k x^{(0)}_{(i)}, \quad k=1, 2, 3, \Lambda, n \quad (1)$$

则生成一次累加生成数列为

$x^{(1)} = \{x^{(1)}_{(1)}, x^{(1)}_{(2)}, \Lambda, x^{(1)}_{(n-1)}, x^{(1)}_{(n)}\}$ 。累加生成使原始无明显规律的离散数据中具有某种特性或规律得以显化, 是灰量积累过程的发展态势^[3]。

(2) 灰色预测建模时, 所使用的有些原始数据, 虽然完整, 但数列可能存在异常数, 剔除异常数则令 $z^{(1)}$ 为数列 $x^{(1)}$ 的紧邻均值生成数列, 即 $z^{(1)} = \{z^{(1)}_{(1)}, z^{(1)}_{(2)}, \Lambda, z^{(1)}_{(n-1)}, z^{(1)}_{(n)}\}$, 其中:

$$z^{(1)}_{(k)} = 0.5x^{(1)}_{(k)} + 0.5x^{(1)}_{(k-1)}, \quad k=2, 3, \Lambda, n \quad (2)$$

(3) GM(1, 1)模型的基本形式建立

$$x^{(0)}_{(k)} + a \cdot z^{(1)}_{(k)} = b \quad (3)$$

其中参数 $-a$ 为发展系数, b 为灰作用量。

引入矩阵向量为:

$$u = \begin{bmatrix} a \\ b \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}_{(2)} \\ x^{(0)}_{(3)} \\ \vdots \\ x^{(0)}_{(n)} \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}_{(2)} & 1 \\ -z^{(1)}_{(3)} & 1 \\ \vdots & \vdots \\ -z^{(1)}_{(n)} & 1 \end{bmatrix} \quad (4)$$

公式(3)可表示为 $Y = Bu$, 采用最小二乘法可求得参数 a, b 的估计值式:

$$\begin{bmatrix} a \\ b \end{bmatrix} = [B^T B]^{-1} B^T Y \quad (5)$$

将 $x^{(0)}_{(k)}$ 的时刻 $k=2, 3, \Lambda, n$ 视为连续变量 t , $\hat{x}^{(1)}_{(t)}$ 视为时间 t 的函数, 则灰方程 GM(1, 1) 的白微分方程(一阶线性非齐次常微分方程):

$$\frac{d\hat{x}^{(1)}_{(t)}}{dt} + a\hat{x}^{(1)}_{(t)} = b \quad (6)$$

(二) GM(1, 1)模型的求解

取初值 $\hat{x}^{(1)}_{(t)}|_{t=1} = x^{(0)}_{(1)}$, 则方程(6)的解为:

$$\hat{x}^{(1)}_{(t)} = \left[x^{(0)}_{(1)} - \frac{b}{a} \right] e^{-at(t-1)} + \frac{b}{a} \quad (7)$$

从而可得方程(3)的时间响应表达式为:

[投稿日期] 2018-01-12

[作者简介] 程帅帅(1992-), 男, 江苏宿迁人, 硕士研究生, 研究方向: 管理科学与工程。

$$\hat{x}_{(k+1)}^{(1)} = \left[x_{(1)}^{(0)} - \frac{b}{a} \right] e^{-ak} + \frac{b}{a}, \quad k=1,2,\Lambda,n-1 \quad (8)$$

对上式作一阶累减生成原始数据列的估值为:

$$\hat{x}_{(k+1)}^{(0)} = \hat{x}_{(k+1)}^{(1)} - \hat{x}_{(k)}^{(1)} = (1-e^a) \left[x_{(1)}^{(0)} - \frac{b}{a} \right] e^{-ak}, \quad k=1,2,\Lambda,n-1 \quad (9)$$

(三) GM(1, 1) 模型精度检验

残差定义如下:

(1) 绝对残差:

$$\varepsilon(k) = x_{(k)}^{(0)} - \hat{x}_{(k)}^{(0)}, \quad k=1,2,3,\Lambda,n \quad (10)$$

(2) 相对残差:

$$\varepsilon_r(k) = \frac{\left| x_{(k)}^{(0)} - \hat{x}_{(k)}^{(0)} \right|}{x_{(k)}^{(0)}} \times 100\%, \quad k=1,2,3,\Lambda,n \quad (11)$$

(3) 平均相对残差:

$$\bar{\varepsilon}_r = \frac{1}{n} \sum_{k=1}^n |\varepsilon_r(k)|, \quad k=1,2,3,\Lambda,n \quad (12)$$

检验预测建模精度, 采用如下精度检验等级参照表 1。

表 1 精度检验等级参照

精度等级	相对残差
一级(好)	1%
二级(合格)	5%
三级(勉强)	10%
四级(不合格)	20%

二、GM(1, 1) 残差尾段修正模型

GM(1, 1) 模型预测的有效性受原始数据的变化和时间预测的长短影响, 如原始数据较少, 建立预测未来长期灰色模型, 模型精度大大降低^[4]; 又或者原始数据过多, 但其可能包含了粗大误差数据, 干扰因素较多, 影响模型精度^[5]。

(一) GM(1, 1) 残差尾段修正模型

当 GM(1, 1) 模型精度不符合要求或者达到特定要求但并不满意时, 此时需要用绝对残差数列建立 GM(1, 1) 模型, 对原模型进行修正, 以提高建模精度, 残差修正模型建立过程如下:

令 $\varepsilon^{(0)} = \{\varepsilon^{(0)}(1), \varepsilon^{(0)}(2), \Lambda, \varepsilon^{(0)}(n)\}$, 其中

$$\varepsilon(k) = x_{(k)}^{(0)} - \hat{x}_{(k)}^{(0)}, \quad k=1,2,3,\Lambda,n \quad (13)$$

其中 $\varepsilon^{(0)}$ 表示绝对残差数列。

若绝对残差数列中的元素存在有正有负, 则需根据式(14)进行正化处理^[6], 得到新数列 $F^{(0)}$:

$$f_{(k)}^{(0)} = \varepsilon_{(k)}^{(0)} + 2|\varepsilon_{(k)\min}^{(0)}|, \quad k=1,2,\Lambda,n \quad (14)$$

上式中, $\varepsilon_{(k)\min}^{(0)}$ 表示 $\varepsilon^{(0)}$ 数列中最小值。

$$F^{(0)} = \{f_{(1)}^{(0)}, f_{(2)}^{(0)}, \Lambda, f_{(k)}^{(0)}\}, \quad k=1,2,\Lambda,n \quad (15)$$

残差修正模型选取绝对残差数列中残差较大的值, 构成残差尾段数列, 根据式(14)正化处理得残差尾段数列 $q^{(0)}$ 。

$$q^{(0)} = \{f_{(k_0)}^{(0)}, f_{(k_0+1)}^{(0)}, \Lambda, f_{(k)}^{(0)}\}, \quad k=1,2,\Lambda,n \quad (16)$$

式(16)残差尾段修正数列按照(1)~(9)式进行求解 GM(1, 1) 模型, 求解得到残差尾段修正模型响应表达式为:

$$q_{(k+1)}^{(1)} = \left(f_{(k_0)}^{(0)} - \frac{b_t}{a_t} \right) e^{-a_t(k-k_0+1)} + \frac{b_t}{a_t}, \quad k=k_0, k_0+1, \Lambda, n \quad (17)$$

则残差尾段修正模型累减还原值为:

$$q_{(k+1)}^{(0)} = q_{(k+1)}^{(1)} - q_{(k)}^{(1)} = (1-e^{a_t}) \left[f_{(k_0)}^{(0)} - \frac{b_t}{a_t} \right] e^{-a_t(k-k_0+1)}, \quad k=k_0, k_0+1, \Lambda, n \quad (18)$$

当 $|a_t|$ 充分小时, $1-e^{a_t} \approx -a_t$, 残差尾段修正模型响应表达式为(导数还原值)^[7]:

$$q_{(k+1)}^{(0)} = (-a_t) \left[f_{(k_0)}^{(0)} - \frac{b_t}{a_t} \right] e^{-a_t(k-k_0+1)}, \quad k=k_0, k_0+1, \Lambda, n \quad (19)$$

则最终灰色预测模型为:

$$x_{(k+1)}^{*(0)} = x_{(k+1)}^{(0)} + \theta_{(k)} (q_{(k+1)}^{(0)} - 2|\varepsilon_{(k)\min}^{(0)}|), \quad k=k_0, k_0+1, \Lambda, n \quad (20)$$

$$\text{其中, } \theta_{(k)} = \begin{cases} 0, & k \leq k_0 \\ 1, & k \geq k_0 \end{cases}.$$

(二) GM(1, 1) 残差尾段修正模型检验

检验步骤如同 1.3 GM(1, 1) 模型精度检验, 同样参照表 1。

三、以建筑业总产值为例

表 2 是我国部分(2008~2017 年)建筑业总产值统计数据, 用残差尾段修正模型模拟后三年建筑业总产值。

表 2 我国建筑业总产值

年份 (序号)	2008 (1)	2009 (2)	2010 (3)	2011 (4)	2012 (5)
建筑业总产值 (亿元)	62036.81	76807.74	96031.13	116463.32	137217.86
年份 (序号)	2013 (6)	2014 (7)	2015 (8)	2016 (9)	2017 (10)
建筑业总产值 (亿元)	160366.06	176713.42	180757.47	193566.78	213954

步骤 1: 基于 GM (1, 1) 模型选取较合适的数据, 作为模型原始值。

(1) 取数列 $x1^{(0)} = \{62036.81, \dots, 193566.78, 213954\}$ 按照 (1) ~ (9) 式计算平均相对残差 $\overline{\varepsilon_r} = 6.78\%$, 其中 2009 年的相对残差最大, 为 23.33%, 用 $x1^{(0)}$ 数据列预测效果不好, 存在误差较大;

(2) 取数列 $x2^{(0)} = \{76807.74, \dots, 193566.78, 213954\}$, 按照 (1) ~ (9) 式计算平均相对残差 $\overline{\varepsilon_r} = 5.15\%$, 其中 2010 年的相对残差较大, 为 15.06%, 用 $x2^{(0)}$ 数据列预测效果不好, 存在误差较大;

(3) 以此类推, $x3^{(0)} = \{96031.13, \dots, 193566.78, 213954\}$, $\overline{\varepsilon_r} = 3.63\%$, 其中 2011 年的相对残差较大, 为 9.30%;

(4) $x4^{(0)} = \{116463.32, \dots, 213954\}$, $\overline{\varepsilon_r} = 2.59\%$, 其中 2012 年的相对残差较大, 为 5.24%, 该年预测精度为三级。

国家统计局公布近 10 年的建筑业总产值, 2011~2017 年数据显示最为完整 (包含了建筑业其他产值), 而 2011 年之前并不包建筑业其他产值, 由上述计算可得近七年 (2011~2017 年) 的平均相对残差最小。

为使我国建筑业总产值预测效果更加准确, 取近七年时间段的数列, 作为残差修正的原始数据, 对此建立残差尾段修正模型。

步骤 2: GM (1, 1) 残差尾段修正模型建立、求解与检验。

(1) 原始数列 $x^{(0)} = x4^{(0)} = \{116463.32, \dots, 193566.78, 213954\}$, 按照式 (1) ~ (9) 得 $\hat{x}_{(k+1)}^{(0)} = 133574.34e^{0.08k}$, 由此可得 GM (1, 1) 模型预测值数列 $\hat{x}^{(0)} = \{116463.32, \dots, 213219.50\}$, 详见表 3;

(2) 绝对残差数列 $\varepsilon^{(0)} = \{0, -7184.30, 4258.36, 7951.30, -1684.87, -3664.72, 734.50\}$;

(3) 则残差尾段 $F^{(0)} = \{-7184.30, 4258.36, 7951.30, -1684.87, -3664.72, 734.50\}$;

(4) 对其按式 (14) 进行正化处理得到 $q^{(0)} = \{7184.30, 18626.95, 22319.89, 12683.73, 10703.88, 15103.09\}$;

(5) 将 $q^{(0)}$ 作为原始数据列按 (17) ~ (19) 式得到 $\hat{q}_{(k+1)}^{(0)} = 22576.22e^{-0.12(k-1)}$;

(6) 根据 (20) 式, 最终得到尾段残差修正模型:

$$\hat{x}_{(k+1)}^{*(0)} = \begin{cases} 133574.34e^{0.08k}, & k \leq 2 \\ 133574.34e^{0.08k} + 22576.22e^{-0.12(k-1)} \\ -14368.60, & k \geq 2 \end{cases}$$

得到最终预测值为: $\hat{x}_{(k+1)}^{*(0)} = \{161708.43, \dots, 211074.86\}$, $k \geq 2$, $\overline{\varepsilon_r} = 1.61\%$, 灰色预测模型精度相对提高, 效果好。详见常规 GM (1, 1) 模型和残差尾段修正模型比较表 3。

表 3 常规 GM (1, 1) 模型和残差尾段修正模型比较

年份 (序号)	实际值 (亿元)	常规 GM (1, 1) 模型		GM (1, 1) 残差尾段修正模型	
		预测值	相对误差 (均值为 2.59%)	预测值	相对误差 (均值为 1.61%)
2011 (1)	116463.32	116463.32	0.00%		
2012 (2)	137217.86	144402.16	5.24%		
2013 (3)	160366.06	156107.70	2.66%	161708.43	0.84%
2014 (4)	176713.42	168762.12	4.50%	172056.97	2.64%
2015 (5)	180757.47	182442.34	0.93%	183697.57	1.63%
2016 (6)	193566.78	197231.50	1.89%	196682.63	1.61%
2017 (7)	213954	213219.50	0.34%	211074.86	1.35%

四、结语

通过 GM (1, 1) 模型和残差尾段修正模型的比较, GM (1, 1) 模型平均相对残差为 2.59%, 基本满足预测要求, 但 2012 年的相对残差为 5.24%, 该年预测精度不高, 达到勉强程度; GM (1, 1) 残差尾段修正模型平均相对残差为 1.61%, 且 2013~2017 年的相对残差为 0.84%, 2.64%, 1.63%, 1.61% 和 1.35%, 各年相对残差不超过 3%, 建模精度提高, 达到预期的预测效果, 结果可靠性提高。GM (1, 1) 残差尾段修正模型预测 2018~2020 年建筑业总产值分别为 226947.37 亿元, 244383.95 亿元和 263479.35 亿元, 当然随着时间的推移, 由于数据过少, 后期模型还需进行残差修正, 并不能长期预测建筑业总产值, 需要不断的修正, 以保持预测的准确性。

参考文献:

[1] 姜慧, 陈晓红. 建筑工程经济[M]. 武汉: 武汉理工大学出

版社, 2014. 8:6.

[2] 李莉敏, 倪稷, 郭巧洪等. 2015 年全国建筑业发展统计分析[J]. 建筑, 2016, (12): 8-23.

[3] 梅正阳, 韩志斌. 数学建模教程[M]. 北京: 科学出版社, 2012. 8: 85-96.

[4] 陈新国, 訾丽. 基于残差修正 GM(1, 1) 模型的年降水量预测研究[J]. 水资源研究, 2014, 133 (4): 4-6.

[5] 柯宏发, 陈永光, 吴金亮. 一种新的基于 GM (1, 1) 模型的粗大误差判别模型[J]. 系统工程与电子技术, 2008, 30 (10): 2003-2006.

[6] 姚裕盛, 徐开俊, 杨泳. 基于 GM(1, 1) 残差修正模型的飞行训练量预测[J]. 西安航空学院学报, 2018, 36 (1): 78-82.

[7] 刘思峰, 杨英杰, 吴利丰等. 灰色系统理论及其应用[M]. 北京: 科学出版社, 2014. 6: 151-155.

[责任编辑 陶爱新]

A research on the forecast of total output of construction industry in China based on GM (1,1) modified residual error model

CHENG Shuai-shuai, SONG Ji-long, LV Hai-tao

(Institute of Management Engineering and Business, Hebei Engineering of University, Handan 056038, China)

Abstract: This paper is focused on the methods for the forecast of total output value of construction industry in China. The total output value of China's construction industry in the past seven years is selected as the research object based on the dynamic grey system GM (1,1) model. The modified residual error model and the correction of residuals' remainder are used to improve the accuracy of the model and forecast results. The simulated analysis indicates that the prediction accuracy of China's construction industry output value has been improved based on GM (1,1) modified residual error model. It could provide reference to the forecast of the total output value of China's short-term construction industry.

Key words: GM(1, 1); modified residual error model; China's total output of construction industry