

基于改进 GM(1,1)模型的中长期人口预测

侯瑞环,徐翔燕

(塔里木大学 信息工程学院,新疆 阿拉尔 843300)

摘要:由于经典 GM(1,1)模型初始值的选取对模型预测精度影响较大,文章选择预测值相邻数据作为初始值,并对此初始值进行参数化,增强模型抗随机干扰能力;同时选择自适应过滤法对预测误差进行修正,增强模型的预测精度。最后,利用改进后的 GM(1,1)模型对新疆人口数量进行预测。结果表明:与经典的 GM(1,1)模型相比,改进后的模型在人口预测中预测精度更高,结果更稳定,更适合中长期人口的预测。

关键词:人口预测;GM(1,1)模型;含参相邻初始值;误差修正

中图分类号:F224.9 文献标识码:A 文章编号:1002-6487(2021)01-0186-03

0 引言

科学地预测人口有助于合理规划城市建设,制定符合实际需求的区域社会经济发展蓝图。人口数量的变化受自然环境、社会环境、文化等多种因素的影响,使得精确预测难度较大,人口预测模型主要有灰色预测模型、回归模型、logistic 模型、马尔萨斯模型、时间序列法、修正指数曲线、人口指数增长模型、神经网络模型等^[1]。由于 GM(1,1)模型作为经典的灰色预测模型,对“少历史数据,信息量不足”的不确定系统预测具有较好的效果,因此 GM(1,1)预测模型得到了广泛的应用。然而,GM(1,1)模型预测对数据序列光滑度、背景值、初始值以及参数的估计方法都有极大的依赖性,李富荣(2013)^[2]通过对原始数据的幂函数变换提升数据光滑度,从而提高模型的预测精度;宋晓华等(2012)^[3]利用滑动平均法对原始数据序列处理,得到了光滑光较高、大致趋势较强的序列,有效地提升了 GM(1,1)模型的预测精度。同时,大量的文献采用牛顿插值法^[4]、高次差值^[5]、lagrange 插值法^[6,7]、组合插值方法^[8-10]、对定积分迭代优化方法^[11]对背景值进行重构,通过调整数据背景值来提升模型的预测精度。在初始值的选取上,传统模型基于旧信息进行预测,谢臣英等(2012)^[12]选取新信息进行数据

模拟,得到的结果精度高于传统 GM(1,1)模型;陈鹏宇等(2019)^[13]认为 GM(1,1)模型所有改进方法中直接求解法最优,但是其优化过程需要设计相应程序来实现,求解难度高过其他改进方法。

本文选取预测值相邻数据作为初始值,在体现新信息优先的同时充分考虑了旧信息的价值,并在初始值中引入参数,增强模型抗随机干扰能力;同时利用自适应过滤法对改进后 GM(1,1)模型预测误差进行修正,以提高模型中长期预测的精度。最后将此模型用于新疆人口数量的中长期预测。

1 GM(1,1)模型的改进

1.1 含参数 δ 的相邻初始值 GM(1,1)模型

经典 GM(1,1)模型基于相加序列进行构建预测方程,假设原始序列:

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\}$$

进行一次累加得到序列 $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)\}$,

则 GM(1,1)模型白化微分方程为:

$$\frac{dX^{(1)}}{dt} + \alpha X^{(1)} = \beta \tag{1}$$

式(1)中, α 、 β 分别为发展灰数与内生控制灰数。

基金项目:塔里木大学校长基金青年创新资金项目(TDZKQN201824)

作者简介:侯瑞环(1986—),男,甘肃会宁人,硕士,讲师,研究方向:统计模型、区域经济。

(通讯作者)徐翔燕(1990—),女,河南罗山人,硕士,讲师,研究方向:智能优化算法、机器学习。

[9]邓聚龙.灰色控制系统[M].武汉:华中理工大学出版社,1993.

[10]汪艳.商贸流通业对制造业全要素生产率影响的实证分析[J].商业经济研究,2017,(21).

[11]魏艳秋,和淑萍,高寿华.“互联网+”信息技术服务业促进制造业升级效率研究——基于 DEA-BCC 模型的实证分析[J].科技管理研究,2018,38(17).

[12]张云凤,王雨.物流产业效率评价及影响因素分析[J].统计与决策,2018,(8).

[13]唐建荣,杜娇娇,唐雨辰.环境规制下的区域物流效率可持续发展研究[J].经济与管理评论,2018,(5).

(责任编辑/方 思)

设 $\gamma=(\alpha, \beta)^T$, 通过最小二乘方法估计得到 $\hat{\gamma}=(A^T A)^{-1} A^T Y$, 其中:

$$A=\begin{pmatrix} -0.5[X^{(0)}(1)+X^{(0)}(2)] & 1 \\ -0.5[X^{(0)}(1)+X^{(0)}(2)] & 1 \\ \vdots & \vdots \\ -0.5[X^{(0)}(1)+X^{(0)}(2)] & 1 \end{pmatrix}$$

$$Y=\left(X^{(0)}(2), X^{(0)}(3), \cdots, X^{(0)}(n)\right)^T$$

求解白化微分方程(1)可以得到预测模型:

$$\hat{X}^{(1)}(k+1)=\left\{X^{(0)}(1)-\frac{\beta}{\alpha}\right\}e^{-\alpha k}+\frac{\beta}{\alpha}, k=1, 2, \cdots, n-1 \quad (2)$$

由于经典的 GM(1,1)模型在预测结果中过分强调 $X^{(0)}(1)$ 所含有的信息,从而导致了 $X^{(0)}(n)$ 所含新信息的损失,反之,则会损失旧信息。为了能够很好地平衡新旧信息的选择,可使用相邻信息作为模型初始值。虽然此方法有效提升了模型的预测精度,但由于预测结果对局部信息的依赖导致了模型的预测周期变短,从而弱化了整体数据趋势信息,使得预测模型只适用于短期预测。本文将参数 δ 引入相邻初始值 $X^{(0)}(k)$, 得到预测方程:

$$\hat{X}^{(1)}(k+1)=\left\{X^{(0)}(k)+\delta-\frac{\beta}{\alpha}\right\}e^{-\alpha}+\frac{\beta}{\alpha}, k=1, 2, \cdots, n-1 \quad (3)$$

参数 δ 通过计算:

$$\min_{\delta} \sum_{k=1}^{n-1} \left\{ \hat{X}^{(1)}(k+1) - X^{(0)}(k+1) \right\}^2$$

得到:

$$\hat{\delta}=\frac{\sum_{k=1}^{n-1} \left\{ \left[X^{(0)}(k+1) - \frac{\beta}{\alpha} \right] + \left[\frac{\beta}{\alpha} - X^{(0)}(k) \right] e^{-\alpha} \right\} e^{-\alpha}}{(n-1)e^{-2\alpha}} \quad (4)$$

则有预测方程:

$$\hat{X}^{(1)}(k+1)=\left\{X^{(0)}(k)+\hat{\delta}-\frac{\beta}{\alpha}\right\}e^{-\alpha}+\frac{\beta}{\alpha}, k=1, 2, \cdots, n-1 \quad (5)$$

累减一次可得:

$$\hat{X}^{(0)}(k+1)=\hat{X}^{(1)}(k+1)-\hat{X}^{(1)}(k) \quad (6)$$

1.2 误差修正

自适应过滤法具备自适应性、需要数据量少、约束条件少等优势,在此使用自适应过滤法对式(6)预测产生的误差序列:

$$e_{k+1}=\hat{X}^{(0)}(k+1)-X^{(0)}(k+1), k=0, 1, \cdots, n-1 \quad (7)$$

进行修正,建立的修正模型:

$$\begin{cases} \hat{e}_{k+1}=\sum_{i=1}^p \hat{\omega}_i e_{k-i+1} \\ \varepsilon_{k+1}=e_{k+1}-\hat{e}_{k+1} \\ \hat{\omega}_i=\omega_i+2\tau \varepsilon_{k+1} e_{k-i+1} \end{cases} \quad (8)$$

其中, $i=1, 2, \cdots, p$, 权重的个数 p 与序列的周期有密切关系,通常取值与序列数据的周期相同,如果序列数

据没有明显的周期性,则用自相关系数。 τ 为学习常数,其取值选择直接影响权数调整的速度,通常情况下,为了能够提升模型(8)逼近最佳权数的速度,选择 $\tau=\frac{1}{p}$, 初始

权重 $\omega_i=\frac{1}{p}, i=1, 2, \cdots, p$, 得到最终预测值为:

$$\hat{X}_{修正}^{(0)}(k+1)=\hat{X}^{(0)}(k+1)+\hat{e}(k+1) \quad (9)$$

2 人口预测分析

2.1 数据说明

选取 2003—2018 年的《新疆统计年鉴》及新疆维吾尔自治区国民经济和社会发展统计公报中新疆人口数量年度数据作为研究序列,数据见表 1。

表 1 新疆 2003—2018 年人口数据 (单位:万人)

年份	2003	2004	2005	2006	2007	2008	2009	2010
人口数量	1933.95	1963.11	2010.35	2050.00	2095.19	2130.81	2158.63	2181.58
年份	2011	2012	2013	2014	2015	2016	2017	2018
人口数量	2208.71	2232.78	2264.30	2298.47	2359.73	2398.08	2444.67	2486.76

2.2 预测模型构建

对表 1 中 2003—2018 年的人口数据分别使用经典的 GM(1,1)、含相邻初始值的 GM(1,1)以及误差修正后含相邻数据初始值的 GM(1,1)构建预测模型对人口数量进行预测,各个模型的预测结果见表 2。

表 2 模型预测结果

年份	经典 GM(1,1)		改进 GM(1,1)		误差修正的改进 GM(1,1)	
	预测值(万人)	相对误差(%)	预测值(万人)	相对误差(%)	预测值(万人)	相对误差(%)
2003	1933.95	0.00	1933.95	0.00	1933.95	0.00
2004	1969.9	0.35	1968.21	0.26	1966.08	0.15
2005	1995.2	0.75	1994.49	0.79	2004.05	0.31
2006	2043.2	0.33	2042.48	0.37	2044.88	0.25
2007	2083.5	0.56	2082.77	0.59	2087.29	0.38
2008	2129.4	0.07	2128.68	0.10	2129.26	0.07
2009	2129.4	1.35	2164.87	0.29	2160.71	0.10
2010	2165.6	0.73	2184.42	0.13	2183.48	0.09
2011	2193.9	0.67	2216.45	0.35	2212.50	0.17
2012	2244.8	0.54	2244.02	0.50	2235.91	0.14
2013	2269.2	0.22	2268.47	0.18	2263.84	0.02
2014	2301.3	0.12	2300.49	0.09	2303.60	0.22
2015	2336.0	1.01	2345.41	1.03	2353.27	0.27
2016	2398.3	0.01	2397.45	0.03	2397.52	0.02
2017	2437.2	0.31	2436.41	0.34	2442.13	0.10
2018	2484.6	0.09	2483.75	0.12	2484.10	0.11
平均相对误差	—	0.47	—	0.34	—	0.16

表 2 结果显示,误差修正后的改进模型预测精度最高,平均相对误差为 0.16%;其次为改进模型,平均相对误差为 0.34%;经典模型预测效果最差,平均相对误差最高为 0.47%。改进后模型预测精度显著高于经典 GM(1,1)预测精度,其预测结果受数据波动影响明显,将此模型用于

中长期预测稳定性较差,而通过自适应过滤法对其进行误差修正后,其预测精度得到提升,同时预测结果相对平稳,适合中长期预测。因此,使用误差修正的改进GM(1,1)模型预测新疆2019—2023年的人口数量,结果如表3所示。

表3 新疆人口数量预测结果 (单位:万人)

年份	2019	2020	2021	2022	2023
预测值	2511.61	2552.18	2593.09	2634.59	2676.72

3 结论

选取合适的初始值对GM(1,1)模型的预测精度至关重要,本文选取预测值相邻数据,同时在初始值中引入参数,较大提升了预测系统的抗干扰能力;并采用自适应过滤法,有效提升了模型的预测精度。相较于经典GM(1,1)模型,本文所构建的改进GM(1,1)模型具有如下优点:

(1)改进GM(1,1)模型既保留了经典GM(1,1)模型计算简单使用容易的优点,同时又有效抵抗了人口预测中随机性因素带来的干扰,确保预测结果比经典模型平均相对误差更小,更能够满足实际预测需求。

(2)自适应过滤法能够在较少约束条件下,通过自动调整权数,提升模型的适应能力。采用自适应过滤法对改进模型的预测误差进行修正,克服了改进模型在中长期预测中的不稳定性,使其能够更好地适用于人口数量的中长期预测。

(3)使用误差修正后的改进GM(1,1)模型对新疆人口数量中长期预测具有较强的预测能力,预测相对误差较低,为人口数量的中长期预测提供一种可行性较高的预测

方法。

参考文献:

- [1]张海峰,杨萍,李春花,等.基于多模型的西宁市人口规模预测[J].干旱区地理,2013,36(5).
- [2]李富荣.改进的动态GM(1,1)模型在人口预测中的应用[J].统计与决策,2013,(19).
- [3]宋晓华,祖丕娥,伊静,等.基于改进GM(1,1)和SVM的长期电量优化组合预测模型[J].中南大学学报(自然科学版),2012,43(5).
- [4]李俊峰,戴文战.基于插值和Newton-Cotes公式的GM(1,1)模型的背景值构建新方法与应用[J].系统工程理论与实践,2004,(10).
- [5]罗贺,胡笑旋,牛艳秋,等.高次插值的GM(1,1)模型预测方法的改进[J].统计与决策,2017,(5).
- [6]唐万梅,向长合.基于二次插值的GM(1,1)模型预测方法的改进[J].中国管理科学,2006,14(6).
- [7]田梓辰,刘森.基于改进灰色GM(1,1)模型的GDP预测实证[J].统计与决策,2018,(11).
- [8]王晓佳,杨善林.基于组合插值的GM(1,1)模型预测方法的改进与应用[J].中国管理科学,2012,20(2).
- [9]李凯,张涛.基于组合插值的GM(1,1)模型背景值的改进[J].计算机应用研究,2018,35(10).
- [10]王璐,沙秀艳,薛颖.改进的GM(1,1)灰色预测模型及其应用[J].统计与决策,2016,(10).
- [11]胡炎丙,陈勇明,赵月.迭代优化背景值的GM(1,1)模型改进[J].统计与决策,2013,(21).
- [12]谢臣英,尹方平.GM(1,1)模型的改进与应用[J].统计与决策,2012,(2).
- [13]陈鹏宇,邓宏伟.GM(1,1)模型的改进现状及应用[J].统计与决策,2019,(6).

(责任编辑/浩 天)