

ARIMA 与指数平滑在区域高等教育规模预测中的比较研究

郑凤霞

(四川文理学院数学与财经学院,四川 达州 635000)

摘要:文章利用在校学生数作为区域高等教育规模指标,建立了基于四川省高等教育在校学生数数据序列的 ARIMA 和指数平滑预测模型,并进行了比较,得出最优模型为 ARIMA(2,1,5) 模型;用此模型对四川省 2013-2015 年高等教育在校学生数做出了预测。结果表明,四川省高等教育在校学生数在近三年内仍会增长,但增长速度趋于缓慢,到 2015 年末,四川省高等教育在校学生数约为 1163100 人。

关键词:区域高等教育规模;ARIMA 模型;指数平滑模型

中图分类号:O212

文献标志码:A

引言

教育是关于民生的重大问题,关系亿万青少年的成长和成才,关系国家民族的前途命运,是当今各国综合国力竞争的战略焦点。

区域高等教育发展规模预测是制定教育规划、教育资源合理配置过程中的重要环节,同时也是政府部门制定教育规划与政策的基础工作,对区域高等教育发展规模进行有效预测具有非常重要的意义。

在高等教育规模预测中,学者们经常用到指数平滑模型,如丁云华^[1]选用线性趋势法、二项式趋势线法、指数趋势线法对我国 2011-2015 年民办高校发展规模进行了预测;佟赤等^[2]用指数平滑模型对成人高等医学教育进行预测;邵云飞等^[3]用趋势拟合模型、二次指数平滑模型、线性回归模型预测四川省研究生教育未来 10 年所应达到的规模;Healey 和 Brown^[4]用指数平滑模型预测大学在校学生数;弗兰克尔和格雷尔德^[5]在进行美国高等教育注册生预测时,根据对入学率进行指数平滑,从而预测出未来的入学率。另外,夏新斌等^[6]用 ARIMA 模型对湖南中等职业教育预测进行了研究。本文建立了基于四川省高等教育在校学生数据序列的

ARIMA 和指数平滑预测模型,并进行了比较,得出最优模型为 ARIMA(2,1,5) 模型;用此模型对四川省 2013-2015 年高等教育在校学生数做出了预测。

1 模型描述

1.1 ARIMA 模型

目前,最常用的用于拟合平稳序列的模型是 ARMA 模型,它的广泛应用源于它的统计性质和著名的 Box-Jenkins 方法^[7]。实际生活中,绝大部分序列都是非平稳的,但是这些序列经过差分后都会显示平稳序列的性质。ARIMA 模型就是由差分运算和 ARMA 模型组合而成。

设自回归模型的阶数为 p , 移动回归模型的阶数为 q , 差分运算的阶数为 d , 则求和自回归移动平均模型记为 $ARIMA(p, d, q)$ ^[8], 其表达式为:

$$\begin{cases} \Phi(B) \nabla^d y_t = \Theta(B) \varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \delta_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(y_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

其中,

$$\begin{cases} \nabla^d = (1 - B)^d \\ \Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p \\ \Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q \end{cases}$$

收稿日期:2013-06-08

基金项目:四川文理学院科研项目(2012Z005Y;2012Z004Z)

作者简介:郑凤霞(1985-),女,湖北恩施人,硕士,主要从事应用数学方面的研究,(E-mail)zhengfengxiade@163.com

1.2 指数平滑模型

平滑法在进行趋势分析的一个常用而又较为简单的方法。移动平均法是用一个简单的加权和作为某一期趋势的预测值,其权重基本相同。但通常实际情况是,远期的结果对现在的影响会小些,近期的结果对现在的影响会大些。指数平滑模型的基本思想就是对不同时期观测值的权重有不同的设置,远期的权重小些,近期的权重大些,各期的权重随时间的间隔的增大而呈指数衰减。指数平滑模型对移动平均法进行了改进和发展,其应用较为广泛。

指数平滑模型分为一次指数平滑模型、二次指数平滑模型和三次指数平滑模型。一次指数平滑模型不能反映趋势变动、季节波动等有规律的变动,它通常只适用于围绕一个稳定水平上下波动的时间序列。所以,对于具有线性趋势的数据,常采用二次指数平滑模型。

设时间序列为 $y_1, y_2, \dots, y_t, \dots$ 则一次指数平滑公式为:

$$S_t^{(1)} = \alpha y_t + (1 - \alpha) S_{t-1}^{(1)}$$

其中, $S_t^{(1)}$ 为第 t 周期的一次指数平滑值; α 为加权系数, $0 < \alpha < 1$ 。于是一次指数平滑预测模型为:

$$\bar{y}_{t+1} = S_t^{(1)} = \alpha y_t + (1 - \alpha) \bar{y}_t$$

即以第 t 周期的一次指数平滑值最为第 $t + 1$ 期的预测值。

设一次指数平滑值为 $S_t^{(1)}$, 则二次指数平滑 $S_t^{(2)}$ 的计算公式为:

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)}$$

于是二次指数平滑预测模型为:

$$\bar{y}_{t+T} = a_t + b_t T$$

其中, \bar{y}_{t+T} 为 $t + T$ 期的预测值; t 为当前期数; T 为预测期数; a_t 为截距, b_t 为斜率,其计算公式为:

$$a_t = 2S_t^{(1)} - S_t^{(2)}$$

$$b_t = \frac{\alpha}{1 - \alpha} (S_t^{(1)} - S_t^{(2)})$$

2 实证分析

2.1 ARIMA 模型

利用高等教育在校学生数作为高等教育规模的指标,根据四川省 1978 - 2011 年高等教育在校学生数进行实证研究,数据来源于四川统计年鉴。首先利用 1978 - 2007 年的 30 个数据建立 ARIMA 模型,然后用所建模型预测 2008 - 2011 年的 4 个数据。

对此数据序列首先取以 10 为底的对数,得到新的数据序列如图 1 所示。

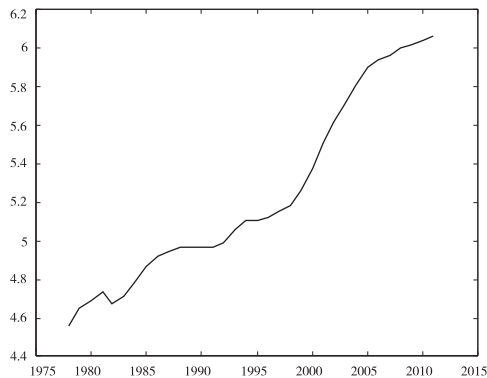


图 1 四川省高等教育在校学生数据序列图

利用 sas9.1.3 软件建立 ARIMA 模型,其建模步骤如下:

(1) 时间序列预处理。根据图 1 知该序列为非平稳序列,又序列蕴含线性趋势,则对序列进行一阶差分运算,差分后的序列为平稳非白噪声序列。

(2) 模型识别。根据差分后序列的自相关系数和偏自相关系数的性质,确定自回归阶数 $p = 2$, 移动回归阶数 $q = 5$ 。

(3) 参数估计。利用条件最小二乘估计 ARIMA 的口径为:

$$(1 + 0.15883B - 0.84116B^2)(1 - B)y_t =$$

$$(1 + 0.89084B + 0.01023B^3 + 0.09014B^5)\varepsilon_t$$

(4) 模型检验。根据 Q_{LB} 统计量的 p 值和 t 统计量的 p 值检验此模型及模型的参数均显著有效。

(5) 序列预测。根据预测方差最小原则确定预测值。预测四川省 2008 - 2011 年高等教育在校学生数见表 1。

表 1 四川省 2008 - 2011 年高等教育在校学生数 ARIMA 预测值

年份	实际值	预测	残差值
2008	5.9961	5.9845	0.0116
2009	6.0153	6.0056	0.0097
2010	6.0359	6.0205	0.0154
2011	6.0566	6.0336	0.0230

2.2 二次指数平滑模型

取平滑系数 $\alpha = 0.2$, 利用 excel2007 软件和截距 a_t 和斜率 b_t 计算公式可得:

$$a_{27} = 2S_{27}^{(1)} - S_{27}^{(2)} = 5.9650$$

$$b_{27} = \frac{0.2}{1 - 0.2} (S_{27}^{(1)} - S_{27}^{(2)}) = 0.0836$$

于是,二次指数平滑预测模型为:

$$\bar{y}_{27+T} = a_{27} + b_{27}T = 5.9650 + 0.0836T$$

预测四川省 2008 - 2011 年高等教育在校学生数见表 2。

表 2 四川省 2008-2011 年高等教育在校学生数指数平滑预测值

年份	实际值	预测值	残差值
2008	5.9961	6.0486	-0.0525
2009	6.0153	6.1322	-0.1169
2010	6.0359	6.2158	-0.1799
2011	6.0566	6.2993	-0.2427

从表 1 和表 2 可以看出 ARIMA 预测法比指数平滑法预测精度要高,所以用 ARIMA 模型对四川省 2013-2015 年高等教育在校学生进行预测,如图 2 所示。用 ARIMA 模型对四川省 2013-2015 年高等教育在校学生进行预测结果见表 3。

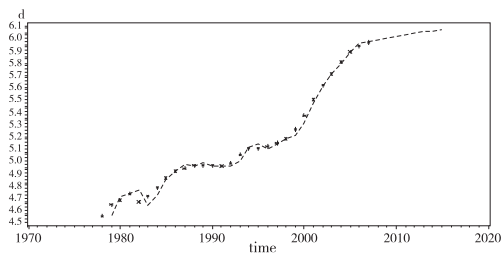


图 2 四川省高等教育在校学生 ARIMA 预测图

表 3 四川省 2013-2015 年高等教育在校学生 ARIMA 预测值

年份	2013	2014	2015
预测值	6.0523	6.0586	6.0656

从图 2 和表 3 可以看出,四川省高等教育在校学生数在未来三年内仍在增长,但增长速度渐趋缓慢。到 2015 年末,四川省高等教育在校学生数为 $10^{6.0656}$, 即约 1 163 100 人。

3 结束语

本文建立了 ARIMA 模型和指数平滑模型,并将其应用于四川省高等教育规模预测中,比较结果表明,ARIMA 模型比指数平滑模型预测精度高。并用 ARIMA 模型预测了四川省未来三年高等教育在校学生数。

参考文献:

- [1] 丁云华.2011-2015 年我国民办高校发展规模预测[J].江西科技学院学报,2012(1):24-29.
- [2] 佟赤,马亚楠,郭军强,等.指数平滑模型在成人高等医学教育发展预测中的应用[J].中国成人教育,2012(8):104-105.
- [3] 邵云飞,赵宏辉,唐小我.四川研究生教育预测模型的实证分析[J].统计与决策,2001(6):69-72.
- [4] Healey H,Marilou T,Brow Daniel G.Forecasting University Enrollments by Ratio Smoothing[J].Higher Education,1978(7):417-429.
- [5] 弗兰克尔 M M,格雷尔德 D E.美国教育统计预测[M].任振华,杨立山,译.北京:高等教育出版社,1988.
- [6] 夏新斌,李琼.ARIMA 模型在教育预测中的应用-以湖南中等职业教育为例[J].经济研究导刊,2008(6):189-190.
- [7] Box G E,Jenkins G.Time Series Analysis,Forecasting and Control[M].San Francisco,CA:Holden-Day,1970.
- [8] 王沁.时间序列分析及其应用[M].成都:西南交通大学出版社,2008.

Comparing Research on ARIMA Model and Exponential Smoothing Model in Regional Higher Education Scale Forecasting

ZHENG Feng-xia

(College of Mathematics and Finance-Economics, Sichuan University of Arts and Science, Dazhou 635000, China)

Abstract: In this paper, the number of students as the scale of regional higher education index, ARIMA and exponential smoothing forecasting model was established based on the number of students in Sichuan higher education. The optimal model was ARIMA(2,1,5) model through comparing two models; Then forecasting the number of students in Sichuan higher education between 2013 and 2015 by ARIMA(2,1,5) model. The result shows that the number of students in Sichuan higher education will still grow in nearly three years, but the growth rate tends to slow, at the end of 2015, the number of students in Sichuan higher education is near about 1163100 people.

Key words: ARIMA model; exponential smoothing model; regional higher education scale