

文章编号:1672-058X(2013)08-0038-07

安徽省城镇化水平预测

——基于IOWHA算子的组合预测*

刘青,杨桂元

(安徽财经大学统计与应用数学学院 安徽 蚌埠 233030)

摘要:以1990—2011年安徽省城镇化水平为样本区间,在指数平滑、灰色预测与回归预测三种单项预测方法的基础上,以预测倒数误差平方和最小为准则,建立IOWHA算子组合预测模型;对该模型的预测有效度的评价结果表明该组合预测模型能有效提高预测精度;根据该组合模型预测出安徽省2012-2015年城镇化水平将分别达到46.35%、47.89%、49.44%、50.98%。

关键词:城镇化水平;IOWHA算子;组合预测

中图分类号:F201

文献标志码:A

0 引言

多年来中国经济保持高速增长,但房地产泡沫、劳动价格上升、货币升值等现状也随之产生,从而使得竞争力出现下降。随着我国竞争力的下降,经济潜在增长率也会受到影响,2012年经济放缓也印证了这一点。当前我国经济面临投资和出口后劲不足等问题,经济结构不合理的深层次矛盾也日益突出,扩大内需,寻求新的驱动成为调整经济结构的首要任务,而最大的内需在城镇化。因此,加快发展城镇化成为促进经济协调发展的必然选择,从而科学地预测城镇化水平具有重要的意义。

我国对城镇化水平的预测主要集中在时间序列预测模型、相关经济因素回归预测、Logistic增长模型、神经网络模型等。丁刚等(2005)用PDL模型预测甘肃省城市化水平将达到30.56%^[1]。任建明等(2006)用PDL模型和Logistic模型对北京市城市化水平进行预测,得出2008年北京城市化水平将达81.89%^[2]。宋丽敏(2007)用简单回归模型、Logistic增长模型、直接预测法与时间序列模型等四种方法分别预测了我国城市化水平,得出到2030年我国城市化水平将达到60%^[3]。然而单项预测方法对信息的提取可能会不够全面,为此本文采用基于IOWHA算子的组合预测模型,综合研究对象多方面的信息,使预测结果更加符合实际。

目前国内外现有的预测方法有很多,不同的预测方法建立在不同的假设前提下,所研究的侧重点也各不相同,在不同的预测方法之间的选择标准,也存在很大的争议,一般总是选择预测误差最小的方法,舍弃其他的方法。自美国学者Yager^[4-6]提出有序加权平均(OWA)算子后,国内外专家学者开始重视有关信息集成算子理论的研究,其中陈华友提出了诱导有序加权算术平均(IOWA)算子和有序加权几何平均(OWGA)

收稿日期:2013-01-10;修回日期:2013-03-07.

* 基金项目:国家社科基金资助项目“组合预测模型与方法创新及其优化理论研究”(12BTJ008).

作者简介:刘青(1987-),女,江西赣州人,从事经济统计研究.

算子,并在此基础上提出了有序加权调和平均(OWHA)算子以及诱导有序加权调和平均算子(IOWHA)算子,建立了新的基于 IOWHA 算子的组合预测模型^[7-10]。

1 模型的选择及建立

自改革开放以来,我国对城镇人口统计的口径不一致^[11],为了尽量减小预测误差,本文只选取第 4 次全国人口普查以后的数据。本文根据 1990—2011 年安徽省城镇人口比重和人均 GDP 数据,选用指数平滑、灰色预测和回归预测建模分别对安徽省城镇化水平做预测。

1.1 单项预测模型

(1) 指数平滑。常见的指数平滑有一次指数平滑、二次指数平滑和三次指数平滑,其中一次指数平滑适用于平稳型时间序列,二次指数平滑适用于具有线性趋势的时间序列,三次指数平滑主要用于具有明显非线性趋势的时间序列。通过对 1990—2011 年安徽省城镇化人口比重的分析,发现该数据近似呈现出线性趋势,如图 1 所示,故本文采用二次指数平滑来预测未来几年安徽省城镇化水平。

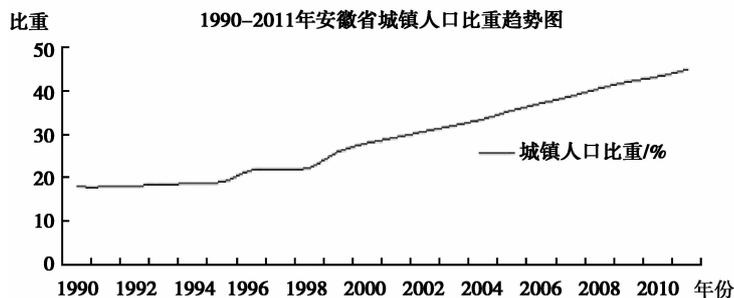


图 1 安徽省城镇人口比重趋势图

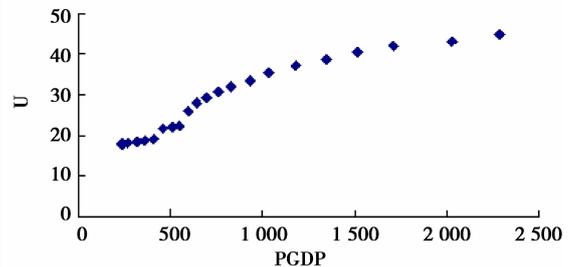


图 2 城镇人口比重与人均 GDP 的散点图

(2) GM(1,1)模型。通过对城镇化人口比重 U 做趋势图(图 1),可看出自 1990 年以来安徽省城镇化水平基本保持着稳定提高的趋势,即城镇化水平随着时间的推进是单调升高的。以往经验表明,GM(1,1)对具有线性趋势的序列做预测能达到较高的预测精度。为此,可利用灰色预测 GM(1,1)模型对城镇人口比重未来的变化趋势做预测。

(3) 回归预测。我国学者周一星曾以城市人口比重代表城市化水平,人均国民生产总值代表经济发展水平,对 1977 年世界 157 个国家和地区的资料进行统计分析,结果表明,绝大多数国家和地区的这两项指标呈现出十分明显的对数曲线相关^[12];夏永祥等曾对我国 1978—1994 年上述两个指标进行处理,得到的散点分布图亦表明我国经济发展水平与城市化水平之间的关系近似于对数曲线相关^[13];胡爱华根据 1978—2002 年我国城市化率与人均 GNP 数据进行回归分析,亦得到了两者之间呈对数相关的研究结果^[14]。张颖等也通过对我国 2000 年各省的城市化率与人均 GDP 数据进行相关分析,发现我国各地区的经济发展水平与城市化水平之间基本符合对数曲线关系^[15,16]。

本文对安徽省 1990—2011 年人均 GDP(PGDP)和城镇人口比重(U)数据做散点图(图 2),发现安徽省城镇化水平与经济发展水平也接近对数关系曲线。于是做出如下回归模型:

$$U_t = -58.491 + 13.382 \ln(\text{PGDP}_t) \\ (-15.339) \quad (23.062) \\ R^2 = 0.964 \quad F = 531.842 \quad SE = 1.816$$

安徽省人均 GDP 的数据是采用 1978 年的不变价,通过指数平滑法预测出 2012—2015 年的人均 GDP 分别为 2 547.64、2 810.93、3 074.22、3 337.51 元,代入上述模型预测出 2011—2015 年城镇人口的比重。

1.2 组合预测模型

传统的组合预测方法是按照单项预测方法的不同而赋予其不同的权重,同一个单项预测方法在其样本区间的各个时点上权重是不变的。但实际上对同一个单项预测方法而言,在不同的时点上其预测精度是不同的,有高有低,即使某一时点上预测较为准确,在另一时点上预测误差较大,因此在组合预测中应依据每个单项预测方法在各个时点的预测精度的高低赋予不同的权重。陈华友(2004)引进诱导有序加权调和平均(IOWHA)算子,提出了以倒数误差平方和为准则新的组合预测模型,克服了现有传统加权组合预测方法赋权的缺陷^[8]。

设 $(v_1, a_1), (v_2, a_2), \dots, (v_m, a_m)$ 为 m 个二维数组,令:

$$f_w[(v_1, a_1), (v_2, a_2), \dots, (v_m, a_m)] = 1 / \sum_{i=1}^m (w_i / a_{v-\text{index}(i)}) \quad (1)$$

则称函数 f_w 是由 v_1, v_2, \dots, v_m 所产生的 m 维诱导有序加权调和平均算子,即 IOWHA 算子。其中 $v-\text{index}(i)$ 是 v_1, v_2, \dots, v_m 中按照从大到小的顺序排序后第 i 个大的数的下标,其中 $W = (w_1, w_2, \dots, w_m)^T$ 是加权向量,满足 $\sum_{i=1}^m w_i = 1, w_i \geq 0, i = 1, 2, \dots, m$ 。从上述概念中可以看出, IOWHA 算子是对诱导值 v_1, v_2, \dots, v_m 按从大到小的顺序排序后所对应的 a_1, a_2, \dots, a_m 中的数进行有序加权调和平均,其中 w_i 与 a_i 的大小和位置无关,而是与其诱导值所在的位置有关。

选择不同预测方法在各个时点上的预测精度作为该方法的诱导值,其中预测精度:

$$v_{it} = \begin{cases} 1 - \left| \frac{(x_t - x_{it})}{x_t} \right| & \left| \frac{(x_t - x_{it})}{x_t} \right| < 1 \\ 0 & \left| \frac{(x_t - x_{it})}{x_t} \right| \geq 1 \end{cases} \quad i = 1, 2, \dots, m; t = 1, 2, \dots, N。$$

v_{it} 表示第 i 种预测方法在第 t 时刻的预测精度,并把 v_{it} 看成预测值 x_{it} 的诱导值。 x_t 为第 t 时刻的实际值, x_{it} 表示第 i 种预测方法在第 t 时刻的预测值。此时, m 种预测方法在 t 时刻的预测精度与其预测值构成了 m 个二维数组: $(v_{1t}, x_{1t}), (v_{2t}, x_{2t}), \dots, (v_{mt}, x_{mt})$ 。

设 $W = (w_1, w_2, \dots, w_m)^T$ 为各种预测方法在组合预测中的加权变量,根据式(1),第 t 时刻的组合预测值:

$$\text{IOWHA}[(v_{1t}, x_{1t}), (v_{2t}, x_{2t}), \dots, (v_{mt}, x_{mt})] = 1 / \sum_{i=1}^m (w_i / x_{v-\text{index}(it)}), t = 1, 2, \dots, N$$

于是, N 期总的基于 IOWHA 的组合预测倒数误差平方和 S^2 :

$$\begin{aligned} S^2 &= \sum_{t=1}^N (1/x_t - 1/\hat{x}_t)^2 = \sum_{t=1}^N \left(\sum_{i=1}^m w_i (1/x_t - 1/x_{v-\text{index}(it)}) \right)^2 \\ &= \sum_{i=1}^m \sum_{j=1}^m w_i w_j \left(\sum_{t=1}^N e_{v-\text{index}(it)} e_{v-\text{index}(jt)} \right) \end{aligned}$$

其中 $e_{v-\text{index}(it)} = 1/x_t - 1/x_{v-\text{index}(it)}$ 。所以,基于倒数误差平方和最小准则的 IOWHA 算子组合预测模型:

$$\begin{aligned} &\min S^2 \\ &\text{s.t.} \begin{cases} \sum_{i=1}^m w_i = 1 \\ w_i \geq 0, i = 1, 2, \dots, m \end{cases} \end{aligned} \quad (2)$$

IOWHA 算子组合预测法通过各单项预测法在各个时点上预测精度的高低按顺序赋权,并以倒数误差平方和最小为准则建立组合预测模型。该组合预测模型具有一定的适用范围,要求样本实际观测值不能太小(如小于1)^[8],本文中城镇人口比重数据符合使用要求,所以本文采用 IOWHA 算子的组合预测模型。

2 实证分析

本文的数据来自《安徽省统计年鉴 2012》,利用 1990—2011 年安徽省城镇人口比重和人均 GDP 作为样本区间,并分别用指数平滑、灰色预测和回归预测作为单项预测法,如表 1 所示。

表 1 实际值与各单项模型预测值、预测精度

年份	实际值 /%	预测值/%			预测精度		
		指数平滑	灰色预测	回归预测	指数平滑	灰色预测	回归预测
1990	17.94	16.19	17.94	15.03	0.902 4	1.000 0	0.837 7
1991	17.96	18.84	15.30	14.58	0.950 8	0.851 7	0.811 8
1992	18.20	19.35	16.62	16.48	0.936 7	0.913 0	0.905 2
1993	18.48	19.39	17.95	18.65	0.950 7	0.971 5	0.990 6
1994	18.78	19.39	19.31	20.36	0.967 4	0.972 0	0.916 0
1995	19.09	19.47	20.67	22.04	0.980 2	0.917 1	0.845 4
1996	21.71	19.63	22.06	23.55	0.904 1	0.984 1	0.915 5
1997	22.02	22.09	23.45	24.95	0.996 9	0.934 9	0.867 0
1998	22.33	22.95	24.87	25.95	0.972 2	0.886 3	0.837 8
1999	26.00	23.26	26.30	27.06	0.894 6	0.988 5	0.959 3
2000	28.00	26.69	27.75	28.03	0.953 2	0.991 0	0.998 8
2001	29.30	29.38	29.21	29.09	0.997 4	0.996 9	0.992 8
2002	30.70	31.02	30.69	30.26	0.989 4	0.999 7	0.985 7
2003	32.00	32.41	32.19	31.42	0.987 1	0.994 1	0.981 9
2004	33.50	33.64	33.70	33.00	0.995 8	0.993 9	0.985 1
2005	35.50	35.03	35.24	34.39	0.986 9	0.992 6	0.968 6
2006	37.10	36.98	36.79	36.15	0.996 9	0.991 5	0.974 4
2007	38.70	38.70	38.35	37.92	0.999 9	0.991 0	0.979 9
2008	40.50	40.33	39.94	39.49	0.995 9	0.986 1	0.975 1
2009	42.10	42.13	41.54	41.11	0.999 3	0.986 7	0.976 4
2010	43.20	43.77	43.16	43.41	0.986 7	0.999 2	0.995 2
2011	44.80	44.88	44.80	45.00	0.998 2	0.999 9	0.995 5

根据式(1),以各时点上各单项预测法的预测精度为诱导值,对各时点预测值进行有序加权平均,并建立以倒数误差平方和最小为准则的式(2),用 Lingo 软件求解,得最优权系数向量为 $W = (0.831, 0.169, 0)^T$,该系数向量表明在 t 时刻组合预测的预测值:

$$\text{IOWHA}[(v_{1t}, x_{1t}), (v_{2t}, x_{2t}), (v_{3t}, x_{3t})] = 1 / (0.831/x_{v-\text{index}(1t)} + 0.169/x_{v-\text{index}(2t)} + 0/x_{v-\text{index}(3t)})$$

对预测精度最高的单项预测法预测值赋权重 0.831,对预测精度次高的单项预测法预测值赋权重 0.169,对预测精度最低的单项预测法预测值赋权重 0,取三者加权调和平均值作为组合预测值。因此根据多种单项预测各时点预测精度的方法来加权组合预测,可以充分利用多种方法间的互补性,对其预测值加权平均后所得组合预测值会更加贴近真实值,能够显著的提高预测精度。表 2 中给出了按预测精度排序后各单项预测方法的预测值、预测精度和组合预测值及其预测精度。

表 2 按预测精度排序后单项预测值

年份	组合预测值/%	组合预测精度	预测值/%			预测精度		
1990	17.62	0.982 1	17.94	16.19	15.03	1.000 0	0.902 4	0.837 7
1991	18.13	0.990 4	18.84	15.30	14.58	0.950 8	0.851 7	0.811 8
1992	18.83	0.965 5	19.35	16.62	16.48	0.936 7	0.913 0	0.905 2
1993	18.53	0.997 2	18.65	17.95	19.39	0.990 6	0.971 5	0.950 7
1994	19.32	0.971 3	19.31	19.39	20.36	0.972 0	0.967 4	0.916 0
1995	19.66	0.970 1	19.47	20.67	22.04	0.980 2	0.917 1	0.845 4
1996	22.29	0.973 1	22.06	23.55	19.63	0.984 1	0.915 5	0.904 1
1997	22.31	0.986 9	22.09	23.45	24.95	0.996 9	0.934 9	0.867 0
1998	23.25	0.958 6	22.95	24.87	25.95	0.972 2	0.886 3	0.837 8
1999	26.42	0.983 7	26.30	27.06	23.26	0.988 5	0.959 3	0.894 6
2000	27.99	0.999 5	28.03	27.75	26.69	0.998 8	0.991 0	0.953 2
2001	29.35	0.998 3	29.38	29.21	29.09	0.997 4	0.996 9	0.992 8
2002	30.75	0.998 5	30.69	31.02	30.26	0.999 7	0.989 4	0.985 7
2003	32.23	0.992 9	32.19	32.41	31.42	0.994 1	0.987 1	0.981 9
2004	33.65	0.995 5	33.64	33.70	33.00	0.995 8	0.993 9	0.985 1
2005	35.20	0.991 6	35.24	35.03	34.39	0.992 6	0.986 9	0.968 6
2006	36.95	0.996 0	36.98	36.79	36.15	0.996 9	0.991 5	0.974 4
2007	38.64	0.998 6	38.70	38.35	37.92	0.999 9	0.991 0	0.979 9
2008	40.27	0.994 3	40.33	39.94	39.49	0.995 9	0.986 1	0.975 1
2009	42.03	0.998 3	42.13	41.54	41.11	0.999 3	0.986 7	0.976 4
2010	43.21	0.999 9	43.16	43.41	43.77	0.999 2	0.995 2	0.986 7
2011	44.82	0.999 6	44.80	44.88	45.00	0.999 9	0.998 2	0.995 5

为能够更为全面地比较 IOWHA 组合预测与各单项预测的预测有效性,本文选择了平方和误差(SSE),均方误差(MAE),平均绝对百分比误差(MAPE),均方根误差(RMSE)4个常用的比较指标,IOWHA 组合预测与各单项预测法的各评价指标值(表 3)。从表 3 中所列各指标值我们看到,IOWHA 组合预测的各项指标表现均优于其他单项预测法,相对于各单项预测方法,IOWHA 组合预测是有效的。

表 3 IOWHA 组合预测的有效性评价

预测效果评价指标体系		SSE	MAE	MAPE	RMSE
单项预测	回归预测	22.434 4	0.651 3	0.029 9	1.009 8
	指数预测	21.367 6	0.677 3	0.029 9	0.985 5
	移动平均	65.955 3	1.326 4	0.059 3	1.731 5
IOWHA 组合预测	W1=0.831				
	W2=0.169	2.859 3	0.264 7	0.011 7	0.360 5
	W3=0				

根据预测的原则,可以用以上求得的组合预测 IOWHA 最优权系数 W 来进行预测区间的 $[N+1, N+2, \dots, N+K]$ 的 IOWHA 组合预测,公式如下:

$$\text{IOWHA}[(v_{1t}, x_{1t}), (v_{2t}, x_{2t}), \dots, (v_{mt}, x_{mt})] = 1 / \sum_{i=1}^m (w_i / x_{v-\text{index}(i)}), t = N+1, N+2, \dots$$

在预测区间 $[N+1, N+2, \dots]$ 上预测精度序列 $v_{1t}, v_{2t}, \dots, v_{mt}$ 的大小确定的原则为根据各单项预测方法在样本区间内近期拟合平均精度的高低,用第 i 种预测方法最近 k 期的拟合平均精度 $\frac{1}{k} \sum_{i=N-k+1}^N v_{ii}$ 来反映预测区间上 $N+k$ 期预测精度的大小^[8]。本文以预测期前四期的平均精度高低作为各单项预测法预测精度高低排序的依据,对 2012—2015 年安徽省城镇化水平做出预测,预测结果如表 4。

表 4 安徽省 2012—2015 年城镇化水平预测值(单位:%)

年份	指数平滑	灰色预测	回归预测	IOWHA 组合预测
2012	46.33	46.46	46.48	46.35
2013	47.84	48.14	47.8	47.89
2014	49.35	49.84	49.01	49.44
2015	50.87	51.56	50.12	50.98

3 结束语

本文利用指数平滑、灰色预测、回归预测三种单项预测法,以 1990—2011 年安徽省城镇化水平为样本区间,建立了以单项预测法预测精度为诱导值,以倒数误差平方和最小为准则的 IOWHA 组合预测模型,得到最优权系数向量为 $W = (0.831, 0.169, 0)^T$,通过对组合预测与单项预测法的各项评价指标值的比较可知,组合预测能够有效的降低单项预测法的预测误差,提高预测的精度。不同的预测方法所侧重的角度是不同的,组合预测法将各单项预测方法进行赋权组合,其综合性更强,故而可以有效的提高预测准确率。

根据 IOWHA 组合预测模型,对各单项预测值进行加权得到 2012—2015 年安徽省城镇化水平将分别达到 46.35%、47.89%、49.44%、50.98%。

参考文献:

- [1] 丁刚,赵萍萍.基于 PDL 模型的城市化水平预测[J].统计研究,2005(3):45-48
- [2] 任建明,孙晖.省域城市化水平预测方法比较——以北京市为例[J].城市问题,2006(3):15-19
- [3] 宋丽敏.中国人口城市化水平预测分析[J].辽宁大学学报,2007,35(3):115-119
- [4] YAGER R. On ordered weighted averaging aggregation operators in multicriteria decision making[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1988, 18(1):183-190.
- [5] YAGER R. Families of OWA operators[J]. Fuzzy Sets and Systems, 1993,59(2):125-148.
- [6] FILEV D, YAGER R. On the issue of obtaining OWA operator weights[J]. Fuzzy Sets and Systems, 1998,94(2):157-169.
- [7] 段妮妮.基于组合预测方法的趋势预测研究与应用[D].西华大学,2009
- [8] 陈华友,刘春林,盛昭瀚.IOWHA 算子及其在组合预测中的应用[J].中国管理科学,2004,12(5):35-40
- [9] 陈华友,刘春林.基于 IOWA 算子的组合预测方法[J].预测,2003,22(6):61-65
- [10] 吴良平,张健,陆媛.基于 IOWHA 算子的组合预测在中国入境旅游中的应用分析[J].旅游学刊,2011,11(26):19-27
- [11] 高宜程,王茂军.区域城镇化水平预测一种方法的讨论——基于晋城市的个案[J].首都师范大学学报,2008,29(1):92-96
- [12] 周一星.城市地理学[M].北京:商务印书馆,1997
- [13] 胡爱华.世界城市化的一般规律和我国的实践[J].经济问题探索,2004(9):115-118
- [14] 夏永祥,余其刚.世界城市化进程的一般规律和中国实践[A].陈甬军,陈爱民.中国城市化:实证分析与对策研究[C].厦门大学出版社,2002
- [15] 张颖,赵民.论区域发展差异与总体城市化水平滞后的关联性[J].城市规划汇刊,2004(4):12-19
- [16] 郭志仪,丁刚.城市化水平预测研究——以 BP 神经网络模型的应用为例[J].人口与经济,2006(6):3-8

Predication for Urbanization Level of Anhui Province ——Combined Forecast Based on IOWHA Operator

LIU Qing, YANG Gui-yuan

(School of Statistics and Applied Mathematics, Anhui University of Finance and Economics,
Anhui Bengbu 233030, China)

Abstract: By taking urbanization level of Anhui Province during 1990-2011 as sample interval, based on such three single forecasts as exponential smoothing, grey forecast and regression forecast, by taking minimum forecast reciprocal error square sum as standard, this paper sets up IOWHA Operator Combination Model, and the evaluation results of forecast effectiveness by this Model show that this combination Model can effectively improve forecast accuracy. Based on this combination Model, this paper forecasts that the urbanization level of Anhui Province during 2012-2015 will reach 46.35 percent, 47.89 percent, 49.44 percent and 50.98 percent respectively.

Key words: urbanization level; IOWHA Operator; combined forecast

责任编辑:代小红