应用概率统计　第二十九卷
第三期　2013年6月

Chinese Journal of Applied Probability
and Statistics Vol.29 No.3 Jun. 2013

成分数据的组合预测 *

张晓琴　陈佳佳　原　静
(山西大学数学科学学院, 太原, 030006)

摘　要

成分数据是一类具有复杂性质的数据, 特别是它的预测研究在管理学与经济学中占有很重要的地位. 组合预测则是近年来在预测中应用比较广泛的一种方法, 它能够充分利用单预测模型的信息, 提高预测精度, 增强预测的稳定性, 且具有较高的适应能力. 本文首次把组合预测方法应用到成分数据的预测分析中, 基于成分数据的一些基本性质, 利用组合预测得到了较好的预测结果.

关键词: 成分数据, 非对称logratio变换, 球坐标变换, 组合预测.

学科分类号: F224.7.

§1. 引　言

成分数据的概念最早来自于1866年Ferrers[1]的工作, 满足

$$S^D=\Big\{x=(x_1,x_2,\ldots,x_D);x_i>0,i=1,2,\ldots,D,\sum_{i=1}^{D}x_i=c\Big\}$$

的空间称之为D维成分数据空间, 这里c是任意常数, S^D中的元素是D维行向量. Aitchison在1986年出版了《The Statistical Analysis of Compositional Data》, 在这本书[2]中, 他的核心思想就是对成分数据做对数比变换, 这样把成分单形空间映射到欧几里得空间中, 从而使得经典的统计方法可以适用于变换后的数据分析, 而后在成分数据的单形空间中, 定义了其上的加法, 乘法, 内积及距离等[3, 4]. 这样就有了成分数据的几何研究, 在此基础上, Egozcue等人提出了等距对数比变换[5], 但是无论是Aitchison提出的对称对数比变换和非对称的对数比变换, 还是后来的等距对数比变换, 对于成分数据中含有零的成分向量就会失去意义, 这是由于零取对数后会出现负无穷的情况. 基于此想法, 王惠文等人提出了多维球坐标变换[6], 这种变换把成分数据映射到多维超球面上, 而不会出现上述问题.

在预测方面, 由于成分数据的特殊性, 对成分数据的预测必须先进行变换, 然后对变换后的数据做预测, 最后将预测结果反解得到预测的成分数据. 成分数据在预测方面也有很多单预测模型: 施久玉和柴艳有把灰色预测应用到成分数据的预测之中[7, 8]; 王惠文等提出成分数据的线性回归模型[9], 并应用回归预测和球坐标变换对成分数据做了有效的预测;

*国家自然科学基金重点项目(71031006)、国家青年基金项目(41101440)和山西省教育厅专项(20120301)资助.
本文2012年10月29日收到, 2012年12月29日收到修改稿.

还可以用时间平滑预测等方法. 组合预测首先于1969年由Bates和Granger提出, 其主要的技术[10]就是把不同的预测技术用于同一事物的预测之中, 综合利用各个预测模型所提供的信息, 可以得到较好的预测结果. 已有研究表明组合预测技术可以使得预测的稳定性增强, 预测精度提高. 最简单的组合预测就是对各个单预测模型结果进行加权组合, 但是权重的选取[11]是许多学者一直研究的问题. 迄今为止, 没有一种求权重的方法可以适用于任何预测模型中, 对权重系数的求解方法还有待进一步地研究. 在衡量组合预测效果的定量标准方面主要有绝对误差, 相对误差, 最小方差等等, 这些标准帮助我们确定组合权重, 进行组合预测. 由于成分数据的特殊性质, 本文将应用已有的单预测模型, 利用组合预测的理论及方法来进一步进行组合预测, 且采用预测值与真实值的Aitchison距离来衡量其预测效果.

本文的结构安排如下: 第2节介绍了成分数据的变换, 其中包括非对称logratio变换和球坐标变换. 第3节介绍组合预测的一些预测标准, 结合成分数据的距离定义, 给出成分数据组合预测求权重的优化模型. 第4节主要做了两个例子, 分别是北京市和中国三次产业的预测分析, 首先对其做变换, 然后对变换后的数据做预测, 最后再经过逆变换把预测数据返回到成分数据作为拟合值. 对两种不同的预测方法做组合得到优于单预测模型的预测结果.

§2. 成分数据的变换

2.1 非对称logratio变换

设$\boldsymbol{x}=(x_1,x_2,\ldots,x_D)$是成分向量, 令

$$y_i=\ln\frac{x_i}{x_D},\qquad i=1,2,\ldots,D-1. \tag{2.1}$$

这种变换称为非对称logratio变换.

如果用y_i表示x_i, 有如下表达式

$$\begin{cases} x_i=e^{y_i}\Big/\Big(1+\sum\limits_{j=1}^{D-1}e^{y_j}\Big), & i=1,2,\ldots,D-1; \\ x_D=1\Big/\Big(1+\sum\limits_{j=1}^{D-1}e^{y_j}\Big). \end{cases}$$

2.2 球坐标变换

对于成分向量$\boldsymbol{x}=(x_1,x_2,\ldots,x_D)$由于定和限制即$x_1+x_2+\cdots+x_D=1$, 首先对成分向量的各分量开根号, $y_i=\sqrt{x_i}$, $i=1,2,\ldots,D$. 此时有$y_1^2+y_2^2+\cdots+y_D^2=1$, 那么向

量$\boldsymbol{y}=(y_1,y_2,\ldots,y_D)$可以看成超球面上的点, 有

$$\begin{cases} y_1=\sin\theta_2\sin\theta_3\sin\theta_4\cdots\sin\theta_D, \\ y_2=\cos\theta_2\sin\theta_3\sin\theta_4\cdots\sin\theta_D, \\ y_3=\cos\theta_3\sin\theta_4\cdots\sin\theta_D, \\ \quad\vdots \\ y_{D-2}=\cos\theta_{D-2}\sin\theta_{D-1}\sin\theta_D, \\ y_{D-1}=\cos\theta_{D-1}\sin\theta_D, \\ y_D=\cos\theta_D. \end{cases}$$

由上式做反变换可得

$$\begin{cases} \theta_D=\arccos y_D, \\ \theta_{D-1}=\arccos\left(\dfrac{y_{D-1}}{\sin\theta_D}\right), \\ \theta_{D-2}=\arccos\left(\dfrac{y_{D-2}}{\sin\theta_D\sin\theta_{D-1}}\right), \\ \quad\vdots \\ \theta_2=\arccos\left(\dfrac{y_2}{\sin\theta_D\sin\theta_{D-1}\cdots\sin\theta_3}\right). \end{cases}$$

我们可以看到球坐标变换把D维的向量$\boldsymbol{y}=(y_1,y_2,\ldots,y_D)\in R^D$映射到超球面$(r,\theta_2,\ldots,\theta_D)\in\Theta^D$, 这里$r^2=\|\boldsymbol{y}\|^2=1$.

§3. 组合预测

3.1 单预测模型介绍

预测技术的发展趋向于多样化, 目前为止, 应用比较广泛的预测模型有回归预测, 时间序列平滑预测, 神经网络预测, 灰色预测等. 时间序列平滑法包括三种预测方法: 移动平均法、加权移动平均法和指数平滑法. 指数平滑法是生产预测中常见的一种方法, 它是移动平均预测方法加以发展的一种特殊加权移动平均预测法, 其原理是任一期的指数平滑值都是本期实际观察值与前一期指数平滑值的加权平均. 根据平滑次数不同, 指数平滑法分为: 一次指数平滑法, 二次指数平滑法, 三次指数平滑法等. 基于后面实例的需要, 本文只介绍三次指数平滑法.

下面给出三次指数平滑预测模型

$$\widehat{x}_{t+T}=a_t+b_tT+c_tT^2,$$

其中a_t, b_t, c_t的计算公式如下

$$\begin{cases} a_t = 3y_t^{(1)} - 3y_t^{(2)} + y_t^{(3)}, \\ b_t = \dfrac{\alpha}{2(1-\alpha)^2}[(6-5\alpha)y_t^{(1)} - (10-8\alpha)y_t^{(2)} + (4-3\alpha)y_t^{(3)}], \\ c_t = \dfrac{\alpha^2}{2(1-\alpha)^2}[y_t^{(1)} - 2y_t^{(2)} + y_t^{(3)}]. \end{cases}$$

$y_t^{(1)}, y_t^{(2)}, y_t^{(3)}$分别表示一次, 二次, 三次指数平滑值, α为权系数.

$$y_t^{(1)} = \alpha x_t + (1-\alpha)y_{t-1}^{(1)}, \quad y_t^{(2)} = \alpha x_t + (1-\alpha)y_{t-1}^{(2)}, \quad y_t^{(3)} = \alpha x_t + (1-\alpha)y_{t-1}^{(3)}.$$

3.2 组合预测的基本原理

组合预测法一般是先用$n\,(n \geq 2)$种不同的单预测模型进行预测, 通过某一个准则来综合多个单预测模型, 从而形成组合模型, 再利用组合模型进行组合预测计算. 本文根据误差平方和最小来求权重, 但对于成分数据而言, 由于成分数据是一个向量, 预测值与真实值之间的误差就不能简单地做差, 而是用成分数据向量之间的距离来作为预测值与真实值之间的误差, 也就是说权重系数是根据预测值与真实值之间的Aitchison距离平方和最小求得的.

Aitchison距离: 设两个成分数据向量: $\boldsymbol{x} = (x_1, x_2, \ldots, x_D)$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_D)$, 则其Aitchison距离为

$$d_a(\boldsymbol{x}, \boldsymbol{y}) = \Big(\sum_{i=1}^{D}\Big(\ln\frac{x_i}{g(\boldsymbol{x})} - \ln\frac{y_i}{g(\boldsymbol{y})}\Big)^2\Big)^{1/2},$$

$g(\boldsymbol{x}) = (x_1 x_2 \cdots x_D)^{1/D}$, $g(\boldsymbol{y}) = (y_1 y_2 \cdots y_D)^{1/D}$.

下面给出求权重系数的步骤:

(1) 假设成分向量$\boldsymbol{x}^t = (x_1^t, x_2^t, \ldots, x_D^t)$有$n\,(n \geq 2)$种预测模型, $\boldsymbol{x}^t$为时间t的真实值. 首先求出第i种预测模型在第t时间的预测值为$\boldsymbol{y}_i^t = (y_{i1}^t, y_{i2}^t, \ldots, y_{iD}^t)$, 预测误差为$d_a(\boldsymbol{x}^t, \boldsymbol{y}_i^t)$. 然后对$n$种单预测模型进行组合, 即

$$\widehat{\boldsymbol{x}}^t = \sum_{i=1}^{n}\beta_i \boldsymbol{y}_i^t, \qquad t = 1, 2, \ldots, N;\ i = 1, 2, \ldots, n,$$

其中, β_i表示第i种预测模型在组合预测中的权重, 且满足$\sum\limits_{i=1}^{n}\beta_i = 1$.

(2) 设SS_e为组合预测模型的Aitchison距离平方和, 即

$$SS_e = \sum_{t=1}^{N} d_a^2(\boldsymbol{x}^t, \widehat{\boldsymbol{x}}^t) = \sum_{t=1}^{N}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_i d_a(\boldsymbol{x}^t, \boldsymbol{y}_i^t)\beta_j d_a(\boldsymbol{x}^t, \boldsymbol{y}_j^t).$$

在约束$\sum\limits_{i=1}^{n}\beta_i = 1$条件下用拉格朗日乘子法最小化$SS_e$即可求出权重$\beta_i$, $i = 1, 2, \ldots, n$.

3.3 预测效果评价

预测作为分析的一种方法, 给决策者以一定依据来做出判断. 那么预测结果的好坏, 直接会影响到决策的正确与否. 评价预测结果好坏的标准我们称之为评价预测的指标. 对于成分数据来讲, 根据成分数据的单形空间中的距离定义, 得到了下面的评价标准. 为了方便说明, 先假设$\boldsymbol{R}$是用本文方法得到的成分数据预测矩阵, $\boldsymbol{X}$是被预测的原始真实数据矩阵, $\boldsymbol{R}, \boldsymbol{X}$都是$N \times D$维矩阵, r_i, x_i分别表示$\boldsymbol{R}, \boldsymbol{X}$的第$i$行, $d_a(x_i, r_i)$表示真实成分数据向量x_i与预测成分数据向量r_i的Aitchison距离. 本文选用如下的预测效果评价指标:

平均Aitchison距离误差(MSD): $\text{MSD} = \Big(\sum\limits_{i=1}^{N} d_a(x_i, r_i)\Big)/N.$

平均Aitchison距离误差越小, 预测效果越好.

§4. 实例分析

4.1 北京市三次产业的预测分析

本例主要对北京市三次产业结构分别做非对称logratio变换与球坐标变换, 然后对变换后的数据进行单预测, 最后进行组合预测, 得到了较好的预测效果.

4.1.1 数据说明

表1 北京市三次产业结构数据及变换后的数据

				非对称logratio变换		球坐标变换	
时间	x_1	x_2	x_3	y_1	y_2	θ_2	θ_3
1991	0.076	0.4868	0.4372	-1.7497	0.1075	-1.7497	-0.4406
1992	0.0686	0.4878	0.4436	-1.8666	0.0950	-1.8666	-0.4666
1993	0.062	0.48	0.458	-1.9997	0.0469	-1.9997	-0.5248
1994	0.069	0.461	0.47	-1.9186	-0.0193	-1.9186	-0.5730
1995	0.058	0.441	0.501	-2.1562	-0.1276	-2.1562	-0.6971
1996	0.052	0.423	0.525	-2.3122	-0.2160	-2.3122	-0.7932
1997	0.047	0.408	0.545	-2.4506	-0.2895	-2.4506	-0.8736
1998	0.043	0.391	0.566	-2.5774	-0.3699	-2.5774	-0.9587
1999	0.04	0.386	0.574	-2.6637	-0.3968	-2.6637	-0.9913
2000	0.037	0.38	0.583	-2.7573	-0.4280	-2.7573	-1.0282
2001	0.0327	0.3622	0.6051	-2.9180	-0.5132	-2.9180	-1.1199
2002	0.0305	0.3475	0.6219	-3.0151	-0.5820	-3.0151	-1.1910
2003	0.0261	0.3581	0.6158	-3.1610	-0.5421	-3.1610	-1.1649
2004	0.024	0.376	0.6	-3.2189	-0.4673	-3.2189	-1.0986
2005	0.0142	0.2943	0.6915	-3.8856	-0.8543	-3.8856	-1.5003

北京市三次产业结构的数据来自统计年鉴. 上面表1给出了北京市(1991–2005年)三次产业结构的成分数据和经过变换后的数据.

表1中, x_1, x_2, x_3分别表示第一产业, 第二产业, 第三产业的比重, 对其做非对称logratio变换后的数据为y_1, y_2, 做球坐标变换后的数据为θ_2, θ_3.

4.1.2 两种变换的回归预测及组合预测

基于表1中的数据, 分别对$y_1, y_2, \theta_2, \theta_3$做回归预测, 然后反变换到成分数据, 得到北京市三次产业结构的拟合值. 下面是两种变换选择的拟合函数.

非对称logratio变换的拟合函数: $\begin{cases} \widehat{y}_1 = -0.0337t - 2.0478, \\ \widehat{y}_2 = -0.0393t + 0.0677. \end{cases}$

球坐标变换的拟合函数: $\begin{cases} \widehat{\theta}_2 = -0.0153t - 0.4216, \\ \widehat{\theta}_3 = -0.0312t + 0.9464. \end{cases}$

基于组合预测的相关理论, 求得$w = (0.45, 0.55)$时组合预测的预测误差平方和最小. 下面的表2, 分别是基于非对称logratio变换的回归预测拟合值, 基于球坐标变化的回归预测拟合值和这两种方法的组合预测拟合值.

表2 各种预测拟合值

时间	非对称logratio变换回归预测			球坐标变换回归预测			组合预测		
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1991	0.0981	0.5302	0.3717	0.0579	0.4777	0.4643	0.0760	0.5013	0.4226
1992	0.0868	0.5110	0.4022	0.0572	0.4688	0.4740	0.0705	0.4878	0.4417
1993	0.0763	0.4907	0.4330	0.0564	0.4600	0.4836	0.0654	0.4738	0.4608
1994	0.0666	0.4693	0.4641	0.0556	0.4511	0.4933	0.0606	0.4593	0.4802
1995	0.0577	0.4470	0.4953	0.0548	0.4422	0.5030	0.0561	0.4444	0.4995
1996	0.0496	0.4238	0.5265	0.0540	0.4333	0.5127	0.0520	0.4290	0.5189
1997	0.0423	0.4000	0.5576	0.0532	0.4245	0.5223	0.0483	0.4135	0.5382
1998	0.0357	0.3757	0.5886	0.0524	0.4156	0.5320	0.0449	0.3976	0.5575
1999	0.0299	0.3510	0.6191	0.0516	0.4068	0.5416	0.0418	0.3817	0.5765
2000	0.0247	0.3261	0.6492	0.0508	0.3981	0.5512	0.0391	0.3657	0.5953
2001	0.0202	0.3011	0.6787	0.0499	0.3893	0.5607	0.0365	0.3496	0.6138
2002	0.0162	0.2762	0.7075	0.0491	0.3807	0.5702	0.0343	0.3337	0.6320
2003	0.0129	0.2516	0.7355	0.0482	0.3721	0.5797	0.0323	0.3179	0.6498
2004	0.0101	0.2273	0.7626	0.0474	0.3635	0.5891	0.0306	0.3022	0.6672
2005	0.0077	0.2036	0.7887	0.0465	0.3550	0.5984	0.0290	0.2869	0.6840

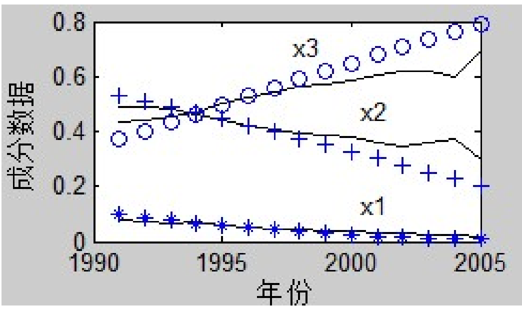

图1　非对称变换回归预测图

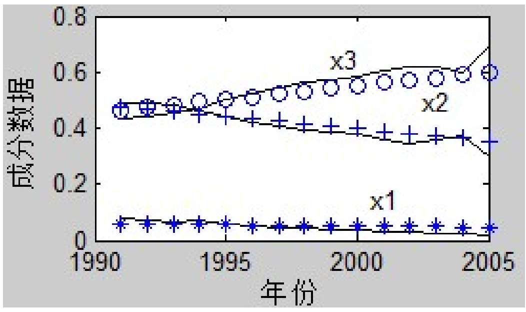

图2　球坐标变换回归预测图

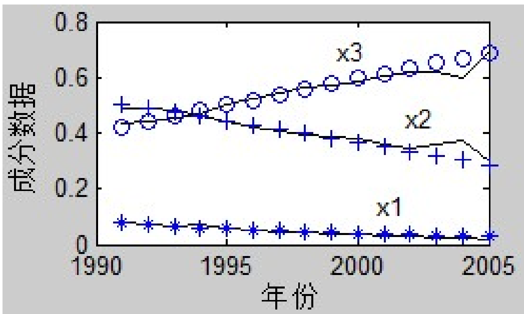

图3　组合预测图

图1, 2, 3给出了各种预测的拟合值与真实值图像表示, 从图像中可以看出组合预测的拟合效果最佳, 图中x1, x2, x3表示原始数据, ‘*’表示x1的拟合数据, ‘+’表示x2的拟合数据, ‘o’表示x3的拟合数据.

表3　各种预测的Aitchison距离误差比较

时间	非对称logratio回归	球坐标回归	组合预测	时间	非对称logratio回归	球坐标回归	组合预测
1991	0.297	0.2454	0.0448	1999	0.2595	0.2243	0.0403
1992	0.2364	0.176	0.0242	2000	0.3618	0.2726	0.0669
1993	0.1915	0.107	0.0483	2001	0.4218	0.3623	0.1045
1994	0.0378	0.1936	0.1146	2002	0.5389	0.4069	0.1132
1995	0.0184	0.0491	0.0301	2003	0.6282	0.5146	0.235
1996	0.0405	0.0456	0.0182	2004	0.7969	0.5771	0.3352
1997	0.0923	0.1177	0.0286	2005	0.5361	0.9796	0.598
1998	0.1616	0.1837	0.0413	MSD	0.3079	0.297	0.1229

表3给出了三种预测方法的拟合值与真实值之间的Aitchison距离, 根据MSD可以看出组合预测优于其他两个单预测模型.

4.2 中国三次产业预测分析

本例对中国三次产业结构做了预测, 运用的单预测模型与前面例子不同, 该例中我们采用了三次平滑预测和回归预测进行组合. 中国三次产业的数来自于统计年鉴. 下面表4列出中国三次产业的成分数据, 及经过非对称logratio变换的数据. 表4中x_1, x_2, x_3分别表示第一产业, 第二产业, 第三产业的百分比重. y_1, y_2表示经过非对称logratio变换的数据. 对表4中变换后的数据做回归预测和三次平滑预测, 得到了如下回归拟合函数.

非对称logratio变换的拟合函数:

$$\begin{cases} \widehat{y}_1 = -0.0017t^2 - 0.07456t - 2.1130, & R_1^2 = 0.9754, \\ \widehat{y}_2 = 0.0077t^2 - 0.1041t + 0.1090, & R_2^2 = 0.8989, \end{cases}$$

其中R_1^2, R_2^2分别表示模型拟合优度指标. 拟合优度指标越接近于1, 模型拟合效果越好.

对表4中变换后的数据y_1, y_2分别用三次平滑法预测, 利用matlab程序得出当$\alpha = 0.16$时对y_1预测均方误差最小, 当$\alpha = 0.27$时对y_2预测均方误差最小. y_1, y_2的拟合值见表5.

表4 原始数据及非对称变换后的数据

时间	x_1	x_2	x_3	y_1	y_2
1995	0.051	0.471	0.478	-2.2378	-0.0148
1996	0.055	0.461	0.484	-2.1748	-0.0487
1997	0.048	0.448	0.504	-2.3514	-0.1178
1998	0.046	0.439	0.515	-2.4155	-0.1597
1999	0.04	0.427	0.533	-2.5896	-0.2217
2000	0.039	0.418	0.543	-2.6335	-0.2616
2001	0.032	0.425	0.543	-2.8314	-0.2450
2002	0.035	0.414	0.551	-2.7564	-0.2859
2003	0.029	0.436	0.535	-2.9150	-0.2046
2004	0.027	0.463	0.51	-2.9386	-0.0967
2005	0.023	0.468	0.509	-3.0970	-0.0840
2006	0.019	0.469	0.512	-3.2939	-0.0877
2007	0.016	0.509	0.475	-3.3907	0.0691

表5 变换后数据经过三次平滑预测的拟合值

时间	y_1	y_2
1995	-2.2378	-0.0148
1996	-2.2076	-0.0423
1997	-2.2718	-0.1109
1998	-2.3467	-0.175
1999	-2.4806	-0.2503
2000	-2.5909	-0.3102
2001	-2.7569	-0.3147
2002	-2.8282	-0.3386
2003	-2.9449	-0.274
2004	-3.0271	-0.1474
2005	-3.1493	-0.0741
2006	-3.3166	-0.0456
2007	-3.4652	0.0889

基于组合预测的相关理论, 求得权重是$w = (1, 0)$时组合预测的预测误差平方和最小. 针对该例子用回归预测会得到好的预测效果, 不需进行组合预测.

§5. 总 结

迄今为止, 对于成分数据的单预测模型研究很多, 但对于成分数据的组合预测的研究几乎没有. 本文是从组合预测的角度来对成分数据做预测分析, 运用实际例子来对三次产业结构进行组合预测. 同时运用Aitchison距离对组合预测误差与单预测误差做了比较. 从文中具体的图和表中都可以看出组合预测的优越性.

把组合预测技术应用到成分数据的预测中, 不仅能够提高成分数据的预测精度, 而且增强了模型的适应性, 对领导阶层科学的判断未来的发展趋势有很大的帮助, 所以说组合预测技术能够给决策者更优的预测结果. 本文首次将组合预测应用到了成分数据的预测当中, 并结合实际例子说明组合预测的优越性.

组合预测技术发展比较迅速, 本文运用的组合预测技术比较简单, 只是运用了两种单预测模型, 求解权重的方法比较简单. 在组合预测中, 权重系数的求解方法有多种, 而对于成分数据的组合预测研究来说, 有没有一种方法能够在统计意义下是最好的, 值得我们做下一步的工作. 前面说的权重都是固定的权值, 那么对于成分数据来说, 有没有一种变权组合即权重随着时间的变化而变化. 而对于组合预测的变权组合在成分数据中又如何应用值得下一步进行研究.

参 考 文 献

[1] Ferrers, N.M., *An Elementary Treatise on Trilinear Coordinates*, London: Macmillan, 1866.

[2] Aitchison, J., *The Statistical Analysis of Compositional Data*, London: Chapman and Hall, 1986.

[3] Aitchison, J. and Egozcue, J.J., Compositional data analysis: where are we and where should we be heading? *Mathematical Geology*, **37(7)**(2005), 829–850.

[4] 张尧庭, *成分数据统计分析引论*, 北京: 科学出版社, 2000.

[5] Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G. and Barceló-Vidal, C., Isometric logratio transformations for compositional data analysis, *Mathematical Geology*, **35(3)**(2003), 279–300.

[6] Wang, H.W., Liu, Q., Mok, H.M.K., Fu, L.H. and Tse, W.M., A hyperspherical transformation forecasting model for compositional data, *European Journal of Operational Research*, **179(2)**(2007), 459–468.

[7] 施久玉, 柴艳有, 灰色成分数据模型在中国产业结构分析预测中的应用, *统计与信息论坛*, **22(1)**(2007), 32–35.

[8] 柴艳有, 灰色成分数据在一类经济结构中的应用, 哈尔滨工程大学硕士学位论文, 2007.

[9] 王惠文, 黄薇, 成分数据的线性回归模型, *系统工程*, **21(2)**(2003), 102–106.

[10] Bates, J.M. and Granger, C.W.J., The combination of forecasts, *Operational Research Quarterly*, **20(4)**(1969), 451–468.

[11] 唐小我, 曹长修, 组合预测方法研究的若干新成果, *预测*, **11(5)**(1992), 39–46.

The Combination Forecast about Compositional Data

Zhang Xiaoqin Chen Jiajia Yuan Jing

(*School of Mathematics Science, Shanxi University, Taiyuan, 030006*)

Compositional data is a kind of complex qualitative data, especially its prediction research plays an important role in management science and economics. The technology of combination forecast is widely used in the forecast, which makes full use of the single forecast model and makes progress on the prediction accuracy. In this paper, the combination forecast method is applied to the prediction of compositional data analysis, based on some basic properties of the compositional data. We can see from the example that using the combination forecast can get a better prediction result.

Keywords: Compositional data, additive logratio transformation, hyperspherical transformation, combination forecast.

AMS Subject Classification: 62G05.