应用概率统计 第二十八卷
第二期 2012年4月

Chinese Journal of Applied Probability
and Statistics Vol.28 No.2 Apr. 2012

基于指数支撑度的最优组合预测模型及其性质研究 *

袁宏俊
(安徽财经大学统计与应用数学学院, 蚌埠, 233030)

陈华友
(安徽大学数学科学学院, 合肥, 230601)

胡凌云
(安徽财经大学管理科学与工程学院, 蚌埠, 233030)

摘 要

在支撑度定义的基础上, 提出平均指数支撑度、平均离散度等概念, 构建了平均指数支撑度最优组合预测模型, 并考虑其等价的平均离散度最优组合预测模型. 针对该模型, 提出优性组合预测等概念, 给出了非劣性组合预测和优性组合预测存在的充分条件、冗余预测方法的存在性和冗余信息的判定等结果, 最后的实例说明了该模型的有效性.

关键词: 组合预测, 平均指数支撑度, 平均离散度, 优性组合预测.

学科分类号: O224.

§1. 引 言

1969年Bates和Granger首次系统提出组合预测方法[1]. 由于组合预测方法能有效地提高预测精度, 增强预测的稳定性, 一直是国内外预测界研究的热点课题[2–10]. 目前理论和应用研究最多的是以误差平方和最小作为最优准则来确定各单项预测方法在组合预测中的加权系数, 唐小我[5]提出优性组合预测等概念, 并利用组合预测绝对误差信息矩阵的性质判断简单算术平均方法是非劣性组合预测、优性组合预测的条件. 马永开[6]提出几个删除冗余方法的定理, 简化非负权重最优组合预测问题。陈华友[7]研究了基于预测有效度的优性组合预测模型及其性质. 王应明[8]提出基于相关性指标的最优组合预测模型, 与传统的组合预测方法有较大的差别. 在此基础上, 陈华友[9]研究了基于向量夹角余弦的相关性的组合预测模型有效性理论. 权双燕[10]提出利用等维递补建立多变量灰色组合预测模型的方法. Yager[11]在提出幂平均算子时, 考虑了数据之间的联系, 提出了支撑度的概念. 本文在支撑度定义的基础上, 建立了平均指数支撑度的最优组合预测模型, 并给出其等价形式的平均离散度的最优组合预测模型. 同时针对该模型, 提出新的优性组合预测、预测方法优超等概念, 探讨非劣性组合预测和优性组合预测存在性、冗余预测方法的存在性及判定方法等性质.

*国家自然科学基金项目(71071002)、国家社会科学基金青年项目(11CJY080)、安徽大学创新团队资助项目(KJTD 001B, SKTD007B)和教育部人文社会科学研究青年基金项目(12YJC630277)资助.

本文2011年2月22日收到.

§2. 基本概念

定义 2.1[11] 设数据$a,b,x,y\in R^+$, 若函数$\mathrm{Sup}(a,b)$满足三个性质:

(1) $\mathrm{Sup}(a,b)\in[0,1]$, 且$\mathrm{Sup}(a,b)=1$当且仅当$a=b$;

(2) $\mathrm{Sup}(a,b)=\mathrm{Sup}(b,a)$;

(3) 如果$|a-b|<|x-y|$, 那么$\mathrm{Sup}(a,b)\geq\mathrm{Sup}(x,y)$,

则称$\mathrm{Sup}(a,b)$为数据b对a的支撑度, 其中R^+为正实数集.

显然, $\mathrm{Sup}(a/(a+b),b/(a+b))=e^{-|a-b|/(a+b)}$满足定义2.1的三个条件, 则称其为数据$b$对$a$归一化的平均指数支撑度, 简记为$\mathrm{Sup}(a,b)$. 如果把数据$a$看成某社会经济现象的指标实际值, b看成其预测值, 则平均指数支撑度在某种程度上也是反映预测精度的一个指标.

设某社会经济现象的指标序列的实际值为$\{x_t,t=1,2,\ldots,N\}$, 设有m种可行的单项预测方法对其进行预测, x_{it}为第i种预测方法在第t时刻的预测值, $i=1,2,\ldots,m$, $t=1,2,\ldots,N$.

定义 2.2 令$\widehat{x}_t=\sum\limits_{i=1}^{m}l_ix_{it}$, 则称$\widehat{x}_t$为实际值$x_t$的组合预测值, 其中$l_1,l_2,\ldots,l_m$为$m$种单项预测在组合预测中的加权系数, 满足$\sum\limits_{i=1}^{m}l_i=1$, $l_i\geq0$, $i=1,2,\ldots,m$.

定义 2.3 令$x_{\max}=\max\{x_t,x_{1t},x_{2t},\ldots,x_{mt}\}$, $x_{\min}=\min\{x_t,x_{1t},x_{2t},\ldots,x_{mt}\}$, 且令

$$y_t=\frac{x_t-x_{\min}}{x_{\max}-x_{\min}},\quad y_{it}=\frac{x_{it}-x_{\min}}{x_{\max}-x_{\min}},\qquad i=1,2,\ldots,m,\quad t=1,2,\ldots,N,$$

则称$\{y_t,t=1,2,\ldots,N\}$为实际值序列$\{x_t,t=1,2,\ldots,N\}$的无量纲归一化序列, $\{y_{it},t=1,2,\ldots,N\}$为第$i$种预测方法对应的预测序列$\{x_{it},t=1,2,\ldots,N\}$的无量纲归一化序列. 显然有$y_t,y_{it}\in[0,1]$.

定义 2.4 令

$$\mathrm{Sup}((y_t,t=1,2,\ldots,N),(y_{it},t=1,2,\ldots,N))=e^{-\sum\limits_{t=1}^{N}\left(|y_t-y_{it}|/\sum\limits_{t=1}^{N}(y_t+y_{it})\right)},$$

则称$\mathrm{Sup}((y_t,t=1,2,\ldots,N),(y_{it},t=1,2,\ldots,N))$为第$i$种预测方法对应的归一化序列和实际值归一化序列的平均指数支撑度.

令$\widehat{y}_t=(\widehat{x}_t-x_{\min})/(x_{\max}-x_{\min})$, 则有

$$\begin{aligned}\widehat{y}_t&=\Big(\sum_{i=1}^{m}l_ix_{it}-x_{\min}\Big)/(x_{\max}-x_{\min})=\Big(\sum_{i=1}^{m}l_i(x_{it}-x_{\min})\Big)/(x_{\max}-x_{\min})\\&=\sum_{i=1}^{m}l_i[(x_{it}-x_{\min})/(x_{\max}-x_{\min})]=\sum_{i=1}^{m}l_iy_{it}.\end{aligned}$$

由定义2.4可得组合预测方法对应的归一化序列和实际值归一化序列的平均指数支撑度为

$$\begin{aligned}&\operatorname{Sup}((y_t,t=1,2,\ldots,N),(\widehat{y}_t,t=1,2,\ldots,N))\\=\ &e^{-\sum\limits_{t=1}^{N}\left(|y_t-\widehat{y}_t|/\sum\limits_{t=1}^{N}(y_t+\widehat{y}_t)\right)}=e^{-\sum\limits_{t=1}^{N}\left(\left|y_t-\sum\limits_{i=1}^{m}l_iy_{it}\right|/\sum\limits_{t=1}^{N}\left(y_t+\sum\limits_{i=1}^{m}l_iy_{it}\right)\right)}.\end{aligned}$$

当组合预测值序列与实际观测值序列完全相同时, 它们对应的归一化序列的平均指数支撑度达到了最大值1. 显然平均指数支撑度越大表示组合预测方法越有效. 因此基于平均指数支撑度的最优组合预测模型可表示为

$$\begin{aligned}&\max\operatorname{Sup}((y_t,t=1,2,\ldots,N),(\widehat{y}_t,t=1,2,\ldots,N))=e^{-\sum\limits_{t=1}^{N}\left(\left|y_t-\sum\limits_{i=1}^{m}l_iy_{it}\right|/\sum\limits_{t=1}^{N}\left(y_t+\sum\limits_{i=1}^{m}l_iy_{it}\right)\right)}\\&\text{s.t.}\ \sum_{i=1}^{m}l_i=1,\ \ l_i\geq 0,\ i=1,2,\ldots,m.\end{aligned}$$

因为指数函数是单调递增的, 所以上述模型等价于如下模型

$$\begin{aligned}&\min f(l_1,l_2,\ldots,l_m)=\sum_{t=1}^{N}\left(\left|y_t-\sum_{i=1}^{m}l_iy_{it}\right|\Big/\sum_{t=1}^{N}\left(y_t+\sum_{i=1}^{m}l_iy_{it}\right)\right)\\&s.t.\ \sum_{i=1}^{m}l_i=1,\ \ l_i\geq 0,\ i=1,2,\ldots,m.\end{aligned}\tag{2.1}$$

定义 2.5 令

$$\begin{aligned}f_i&=\sum_{t=1}^{N}\left(\left|y_t-y_{it}\right|\Big/\sum_{t=1}^{N}(y_t+y_{it})\right),\\f(l_1,l_2,\ldots,l_m)&=\sum_{t=1}^{N}\left(\left|y_t-\sum_{i=1}^{m}l_iy_{it}\right|\Big/\sum_{t=1}^{N}\left(y_t+\sum_{i=1}^{m}l_iy_{it}\right)\right),\end{aligned}$$

则称f_i为第i种预测方法对应的归一化序列和实际值归一化序列的平均离散度. 称$f(l_1,l_2,\ldots,l_m)$为组合预测法对应的归一化序列和实际值归一化序列的平均离散度.

记$f_{\min}=\min\limits_{1\leq i\leq m}\{f_i\}$, $f_{\max}=\max\limits_{1\leq i\leq m}\{f_i\}$, $f_{\min}$和$f_{\max}$分别表示最小的平均离散度和最大的平均离散度.

定义 2.6 若$f(l_1,l_2,\ldots,l_m)>f_{\max}$, 则称权系数$l_1,l_2,\ldots,l_m$确定的组合预测模型为劣性组合预测, 若$f_{\max}\geq f(l_1,l_2,\ldots,l_m)\geq f_{\min}$, 则称之为非劣性组合预测, 若$f(l_1,l_2,\ldots,l_m)<f_{\min}$, 则称之为优性组合预测.

定义2.6表明优性组合预测对应的平均离散度要比最好的单项预测方法对应的平均离散度还要小.

定义 2.7 若第i种单项预测方法、第j种单项预测方法的无量纲归一化指标满足

$$|y_t-y_{it}|\leq|y_t-y_{jt}|\ \text{且}\ y_{it}\geq y_{jt},\qquad t=1,2,\ldots,N,$$

则称基于平均离散度意义下第i种单项预测方法优超第j种单项预测方法, 若对任意时刻均有严格的不等号成立, 则称第i种单项预测方法严格优超第j种单项预测方法.

定义 2.8　若某种单项预测方法添加到组合预测模型中不能使组合预测的平均离散度减少, 则称该单项预测方法为冗余预测方法, 即在最优组合预测模型中的权系数为零. 在一个组合预测模型中, 设有m种单项预测方法参与组合预测, 若最优解中出现冗余方法的个数为m', 则称比例系数$\mu = m'/m$为组合预测模型的冗余度.

显然$0 \le \mu \le (m-1)/m$, 冗余度越小表示组合预测模型选择的单项预测方法越有效, $\mu = 0$说明所有的单项预测方法在同一个预测模型中所提供的信息都是有效的.

§3.　性质定理

定理 3.1　基于平均离散度组合预测模型(2.1)的任一满足约束条件的解对应的组合预测至少是非劣性组合预测.

证明:　设$l_1, l_2, \ldots, l_m$为组合预测模型(2.1)的任一满足约束条件的解, 则有

$$\sum_{i=1}^{m} l_i = 1, \qquad l_i \ge 0,\ i = 1, 2, \ldots, m,$$

该可行解对应的组合预测模型的平均离散度为

$$f(l_1, l_2, \ldots, l_m) = \sum_{t=1}^{N} \left(\left| y_t - \sum_{i=1}^{m} l_i y_{it} \right| \Big/ \sum_{t=1}^{N} \left(y_t + \sum_{i=1}^{m} l_i y_{it} \right) \right). \tag{3.1}$$

由于$\left| \sum\limits_{i=1}^{m} l_i (y_t - y_{it}) \right| \le \sum\limits_{i=1}^{m} l_i |y_t - y_{it}|$, $\sum\limits_{i=1}^{m} l_i = 1$, 所以有

$$\begin{aligned} f(l_1, l_2, \ldots, l_m) &= \sum_{t=1}^{N} \left| \sum_{i=1}^{m} l_i (y_t - y_{it}) \right| \Big/ \sum_{t=1}^{N} \left(y_t + \sum_{i=1}^{m} l_i y_{it} \right) \\ &\le \sum_{i=1}^{m} l_i \left(\sum_{t=1}^{N} |y_t - y_{it}| \right) \Big/ \sum_{i=1}^{m} l_i \left(\sum_{t=1}^{N} (y_t + y_{it}) \right). \end{aligned} \tag{3.2}$$

因为$f_i = \sum\limits_{t=1}^{N} |y_t - y_{it}| \Big/ \sum\limits_{t=1}^{N} (y_t + y_{it})$, 则$\sum\limits_{t=1}^{N} |y_t - y_{it}| = f_i \sum\limits_{t=1}^{N} (y_t + y_{it})$, 代入(3.2)得

$$\begin{aligned} f(l_1, l_2, \ldots, l_m) &\le \sum_{i=1}^{m} l_i f_i \left(\sum_{t=1}^{N} (y_t + y_{it}) \right) \Big/ \sum_{i=1}^{m} l_i \left(\sum_{t=1}^{N} (y_t + y_{it}) \right) \\ &\le \sum_{i=1}^{m} l_i f_{\max} \left(\sum_{t=1}^{N} (y_t + y_{it}) \right) \Big/ \sum_{i=1}^{m} l_i \left(\sum_{t=1}^{N} (y_t + y_{it}) \right) = f_{\max}. \end{aligned}$$

由定义2.6得证.　　□

推论 3.1　在平均离散度组合预测模型(2.1)中, 简单平均组合预测方法一定是非劣性的.

定理 3.2 假设$f_s=f_{\min}$, 即f_s表示m种单项预测方法中的最小平均离散度, 且存在非负加权系数$l_1,l_2,\ldots,l_m$, 使得$y_t\geq y_{st}$, $y_t-\sum\limits_{i=1}^m l_iy_{it}\in(y_{st}-y_t,y_t-y_{st})$, $t=1,2,\ldots,N$, 则基于平均离散度的组合预测模型(2.1)一定存在优性组合预测.

证明: 因为存在非负加权系数$l_1,l_2,\ldots,l_m$, 使得$y_t\geq y_{st}$, $y_t-\sum\limits_{i=1}^m l_iy_{it}\in(y_{st}-y_t,y_t-y_{st})$, $t=1,2,\ldots,N$, 则有

$$y_t-\sum_{i=1}^m l_iy_{it}<y_t-y_{st},\qquad \text{即}\sum_{i=1}^m l_iy_{it}>y_{st}.$$

代入(3.1)中有

$$\begin{aligned}f(l_1,l_2,\ldots,l_m)&=\sum_{t=1}^N\Big|y_t-\sum_{i=1}^m l_iy_{it}\Big|\Big/\sum_{t=1}^N\Big(y_t+\sum_{i=1}^m l_iy_{it}\Big)\\&<\sum_{t=1}^N|y_t-y_{st}|\Big/\sum_{t=1}^N(y_t+y_{st})=f_s=f_{\min}.\end{aligned}$$

由定义2.6得证. □

推论 3.2 假设$f_s=f_{\min}$, 即f_s表示m种单项预测方法中的最小平均离散度, 且

$$y_t\geq y_{st},\quad y_t-\frac{1}{m}\sum_{i=1}^m y_{it}\in(y_{st}-y_t,y_t-y_{st}),\qquad t=1,2,\ldots,N,$$

则简单平均组合预测方法一定是优性组合预测.

定理 3.3 假设基于平均离散度的组合预测模型(2.1)的冗余度$\mu<(m-1)/m$, 则其最优解对应的组合预测一定为优性组合预测方法.

证明: 设基于平均离散度的组合预测模型(2.1)的最优解为$l_1^*,l_2^*,\ldots,l_m^*$, 而$l_1,l_2,\ldots,l_m$为其任一可行解而不是最优解, 由于平均离散度的组合预测模型(2.1)目标函数是求最小值, 则有

$$f(l_1^*,l_2^*,\ldots,l_m^*)<f(l_1,l_2,\ldots,l_m),\tag{3.3}$$

$\sum\limits_{i=1}^m l_i^*=1$, $l_i^*\geq0$, $\sum\limits_{i=1}^m l_i=1$, $l_i\geq0$, $i=1,2,\ldots,m$. 根据定义2.5可知$f(1,0,\ldots,0)=f_1$, $f(0,1,\ldots,0)=f_2,\ldots,f(0,0,\ldots,1)=f_m$. 因为组合预测模型(2.1)的冗余度$\mu<(m-1)/m$, 所以最优解中至少有两个非零分量, 而$(1,0,\ldots,0)^T,(0,1,\ldots,0)^T,\ldots,(0,0,\ldots,1)^T$中只有一个非零分量, 所以它们均是组合预测模型(2.1)的可行解向量而不是最优解向量, 由(3.3)得

$$\begin{cases}f(l_1^*,l_2^*,\ldots,l_m^*)<f(1,0,\ldots,0)=f_1\\f(l_1^*,l_2^*,\ldots,l_m^*)<f(0,1,\ldots,0)=f_2\\\qquad\cdots\\f(l_1^*,l_2^*,\ldots,l_m^*)<f(0,0,\ldots,1)=f_m\end{cases}.$$

所以

$$f(l_1^*, l_2^*, \ldots, l_m^*) < \min_{1 \le i \le m} \{f_i\} = f_{\min},$$

由定义2.6得证. $\square$

定理 3.4 设$f(l_1^*, l_2^*, \ldots, l_m^*)$和$f(\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_m, \bar{l}_{m+1})$分别表示$m$种单项预测方法及$m+1$种单项预测方法参与的组合预测模型(2.1)对应的最小平均离散度, 则

$$f(\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_m, \bar{l}_{m+1}) \le f(l_1^*, l_2^*, \ldots, l_m^*),$$

即基于平均离散度组合预测模型(2.1)的最优目标函数值是m的单调不增函数.

证明: 设$l_1^*, l_2^*, \ldots, l_m^*$为$m$种单项预测方法参与的组合预测模型(2.1)的最优解, $\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_m, \bar{l}_{m+1}$为$m+1$种单项预测方法参与的组合预测模型(2.1)的最优解, 则

$$f(l_1^*, l_2^*, \ldots, l_m^*) = \sum_{t=1}^{N} \Big(\Big| y_t - \sum_{i=1}^{m} l_i^* y_{it} \Big| \Big/ \sum_{t=1}^{N} \Big(y_t + \sum_{i=1}^{m} l_i^* y_{it} \Big) \Big), \tag{3.4}$$

$$f(\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_{m+1}) = \sum_{t=1}^{N} \Big(\Big| y_t - \sum_{i=1}^{m+1} \bar{l}_i y_{it} \Big| \Big/ \sum_{t=1}^{N} \Big(y_t + \sum_{i=1}^{m+1} \bar{l}_i y_{it} \Big) \Big),$$

其中$\sum\limits_{i=1}^{m} l_i^* = 1$, $l_i^* \ge 0$, $i = 1, 2, \ldots, m$; $\sum\limits_{i=1}^{m+1} \bar{l}_i = 1$, $\bar{l}_i \ge 0$, $i = 1, 2, \ldots, m+1$. 显然$(l_1^*, l_2^*, \ldots, l_m^*, 0)^T$为$m+1$种单项预测方法的组合预测模型(2.1)的可行解向量, 由于平均离散度的组合预测模型(2.1)目标函数是求最小值, 则有

$$\begin{aligned} f(\bar{l}_1, \bar{l}_2, \ldots, \bar{l}_{m+1}) &\le f(l_1^*, l_2^*, \ldots, l_m^*, 0) \\ &= \sum_{t=1}^{N} \Big(\Big| y_t - \sum_{i=1}^{m} l_i^* y_{it} \Big| \Big/ \sum_{t=1}^{N} \Big(y_t + \sum_{i=1}^{m} l_i^* y_{it} \Big) \Big) \\ &= f(l_1^*, l_2^*, \ldots, l_m^*). \end{aligned}$$

故得证. $\square$

定理3.4表明在实际过程中组合预测模型(2.1)再增加一种单项预测方法时, 对应的最小平均离散度可能不变, 表明在一定条件下组合预测模型(2.1)可能存在冗余预测方法.

定理 3.5 假设第j种单项预测方法严格优超第k种单项预测方法, 且组合预测的无量纲归一化指标满足: $y_t \ge y_{it}$或$y_t \le y_{it}$, $i = 1, 2, \ldots, m$, $t = 1, 2, \ldots, N$, 则组合预测模型(2.1)的冗余度至少为$1/m$.

证明: (反证法)假设组合预测模型(2.1)的冗余度为0, 不妨设$L^* = (l_1^*, \ldots, l_j^*, \ldots, l_k^*, \ldots, l_m^*)^T$为组合预测模型(2.1)的最优解向量, 其中$l_i^* > 0$, $\sum\limits_{i=1}^{m} l_i^* = 1$, $i = 1, 2, \ldots, m$. 因为$y_t \ge y_{it}$或$y_t \le y_{it}$, $i = 1, 2, \ldots, m$, $t = 1, 2, \ldots, N$, 由(3.4)最优解L^*对应的平均离散度为

$$\begin{aligned} f(l_1^*, l_2^*, \ldots, l_m^*) &= \sum_{t=1}^{N} \Big| \sum_{i=1}^{m} l_i^* (y_t - y_{it}) \Big| \Big/ \sum_{t=1}^{N} \Big(y_t + \sum_{i=1}^{m} l_i^* y_{it} \Big) \\ &= \sum_{t=1}^{N} \sum_{i=1}^{m} l_i^* |y_t - y_{it}| \Big/ \sum_{t=1}^{N} \Big(y_t + \sum_{i=1}^{m} l_i^* y_{it} \Big). \end{aligned}$$

构造$\overline{L}=(l_1^*,\ldots,l_j^*+l_k^*,\ldots,0,\ldots,l_m^*)^T$, 显然$\overline{L}$为组合预测模型(2.1)的一个可行解向量, 其对应的平均离散度为

$$f(l_1^*,\ldots,l_j^*+l_k^*,\ldots,0,\ldots,l_m^*)=\frac{\sum\limits_{t=1}^{N}\Big(\sum\limits_{i\neq j,k}l_i^*|y_t-y_{it}|+(l_j^*+l_k^*)|y_t-y_{jt}|\Big)}{\sum\limits_{t=1}^{N}\Big(y_t+\sum\limits_{i\neq j,k}l_i^*y_{it}+(l_j^*+l_k^*)y_{jt}\Big)}. \tag{3.5}$$

因为第j种预测方法严格优超第k种预测方法, 即$|y_t-y_{jt}|<|y_t-y_{kt}|$且$y_{jt}>y_{kt}$, 代入(3.5)有

$$f(\overline{L})<\frac{\sum\limits_{t=1}^{N}\Big(\sum\limits_{i\neq j,k}l_i^*|y_t-y_{it}|+l_j^*|y_t-y_{jt}|+l_k^*|y_t-y_{kt}|\Big)}{\sum\limits_{t=1}^{N}\Big(y_t+\sum\limits_{i\neq j,k}l_i^*y_{it}+l_j^*y_{jt}+l_k^*y_{kt}\Big)}=\frac{\sum\limits_{t=1}^{N}\sum\limits_{i=1}^{m}l_i^*|y_t-y_{it}|}{\sum\limits_{t=1}^{N}\Big(y_t+\sum\limits_{i=1}^{m}l_i^*y_{it}\Big)}=f(L^*).$$

这与$L^*=(l_1^*,\ldots,l_j^*,\ldots,l_k^*,\ldots,l_m^*)^T$为组合预测模型(2.1)的最优解向量矛盾! 所以假设不成立. □

定理 3.6 假设组合预测中某一种单项预测方法优超其余任一种单项预测方法, 且组合预测的无量纲归一化指标满足: $y_t\geq y_{it}$或$y_t\leq y_{it}$, $i=1,2,\ldots,m$, $t=1,2,\ldots,N$, 则组合预测模型(2.1)的冗余度为$(m-1)/m$.

证明: 假设第p种单项预测方法都优超其余任一种单项预测方法, 即有

$$|y_t-y_{pt}|\leq|y_t-y_{it}|\ \text{且}\ y_{pt}\geq y_{it},\qquad i=1,2,\ldots,m\ \text{且}\ i\neq p,\ t=1,2,\ldots,N. \tag{3.6}$$

由定义2.5知$f_p\leq f_i$, $i=1,2,\ldots,m$且$i\neq p$, 即$f_p=f_{\min}$. 因为$y_t\geq y_{it}$或$y_t\leq y_{it}$, $i=1,2,\ldots,m$, $t=1,2,\ldots,N$, 注意到$\sum\limits_{i=1}^{m}l_i=1$, 则有

$$\sum_{t=1}^{N}\Big|y_t-\sum_{i=1}^{m}l_iy_{it}\Big|=\sum_{t=1}^{N}\Big|\sum_{i=1}^{m}l_i(y_t-y_{it})\Big|=\sum_{t=1}^{N}\sum_{i=1}^{m}l_i|y_t-y_{it}|. \tag{3.7}$$

由(3.1)、(3.6)和(3.7), 模型(2.1)的任一可行解为$l_1,l_2,\ldots,l_m$对应的平均离散度为

$$\begin{aligned}f(l_1,l_2,\ldots,l_m)&=\sum_{t=1}^{N}\sum_{i=1}^{m}l_i|y_t-y_{it}|\Big/\sum_{t=1}^{N}\Big(y_t+\sum_{i=1}^{m}l_iy_{it}\Big)\\&\geq\sum_{t=1}^{N}\sum_{i=1}^{m}l_i|y_t-y_{pt}|\Big/\sum_{t=1}^{N}\Big(y_t+\sum_{i=1}^{m}l_iy_{pt}\Big)\\&=\sum_{t=1}^{N}|y_t-y_{pt}|\Big/\sum_{t=1}^{N}(y_t+y_{pt})=f_p=f_{\min}.\end{aligned}$$

即组合预测模型(2.1)的最优解仅为$L^*=(l_1^*,\ldots,l_p^*,\ldots,l_m^*)^T=(0,\ldots,1,\ldots,0)^T$, 所以冗余度为$(m-1)/m$. □

§4.　实证分析

以职工平均工资收入预测为例说明本文提出的基于平均离散度最优组合预测模型的有效性, 职工平均工资水平是衡量一个地区经济和社会发展水平的重要指标之一. 应用某省1999–2008年职工平均工资实际值(x_t)的数据[12], 分别利用灰色预测模型(x_{1t})、指数平滑模型(x_{2t})、二次曲线函数模型(x_{3t})三种单项预测方法为基础构建组合预测模型, 基础数据如下

表1　职工平均工资收入实际值和三种单项预测方法预测值(单位: 元)

年份	实际值x_t	预测值x_{1t}	预测值x_{2t}	预测值x_{3t}
1999	6516	6516.0	6921.2	6081.0
2000	6989	6825.9	7292.5	7323.1
2001	7908	7595.2	8094.7	8749.0
2002	9296	8940.7	9389.0	10358.8
2003	10581	10840.2	10352.3	12152.3
2004	12928	13608.7	11035.7	14129.7
2005	15334	15637.9	13119.6	16290.9
2006	17949	18248.5	17821.0	18635.9
2007	22180	21671.4	21405.1	21164.7
2008	26363	25380.2	25902.8	23877.3

按定义2.3对表1数据进行无量纲归一化处理见表2如下

表2　实际值和三种单项预测方法预测值的无量纲归一化指标

年份	实际值y_t	预测值y_{1t}	预测值y_{2t}	预测值y_{3t}
1999	0.5177	0.5177	1	0
2000	0.3280	0	0.9385	1
2001	0.2711	0	0.4329	1
2002	0.2505	0	0.3161	1
2003	0.1271	0.2711	0	1
2004	0.6116	0.8316	0	1
2005	0.6983	0.7941	0	1
2006	0.1571	0.5246	0	1
2007	1	0.4991	0.2368	0
2008	1	0.6046	0.8149	0

代入到组合预测模型(2.1)中, 利用LINGO软件计算得最优权系数为: $l_1 = 0.6549$, $l_2 = 0.2781$, $l_3 = 0.067$, 从而可得三种单项预测方法的组合预测值$\widehat{x}_t$如下

表3 指标实际值和基于平均离散度的最优组合预测值

年份	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
x_t	6516	6989	7908	9296	10581	12928	15334	17949	22180	26363
$\widehat{x}_t$	6599.5	6989.0	7811.4	9160.4	10792.4	12928.1	14981.3	18155.6	21563.4	25424.8

按照组合预测效果评价的原则, 采用下列误差指标作为评价指标体系:

(1) 平方和误差

$$\mathrm{SSE} = \sum_{t=1}^{N}(x_t - \widehat{x}_t)^2;$$

(2) 均方误差

$$\mathrm{MSE} = \frac{1}{N}\sqrt{\sum_{t=1}^{N}(x_t - \widehat{x}_t)^2};$$

(3) 平均绝对误差

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}|x_t - \widehat{x}_t|;$$

(4) 平均绝对百分比误差

$$\mathrm{MAPE} = \frac{1}{N}\sum_{t=1}^{N}|(x_t - \widehat{x}_t)/x_t|;$$

(5) 均方百分比误差

$$\mathrm{MSPE} = \frac{1}{N}\sqrt{\sum_{t=1}^{N}[(x_t - \widehat{x}_t)/x_t]^2}.$$

表4 各种预测方法预测效果评价指标体系

	SSE	MSE	MAE	MAPE	MSPE
灰色预测方法	2187846	147.9	386.6	0.0275	0.0098
指数平滑预测方法	9665114	310.9	668.7	0.0511	0.0225
二次函数预测方法	14647767	382.7	1059.1	0.0817	0.0280
基于平均离散度的组合预测方法	1506806	122.7	264.1	0.0157	0.0060
基于误差平方和最小的组合预测方法 ($l_1 = 0.7568$, $l_2 = 0.2432$, $l_3 = 0$)	1326611	115.2	271.4	0.0180	0.0064

上表给出了各种预测方法预测效果评价结果, 可以看出, 基于平均离散度的组合预测模型的各误差指标均低于三种单项预测模型预测误差指标值, 表明提出的组合预测方法能够有效提高预测精度. 同时, 基于平均离散度组合预测模型的各误差指标多数优于传统的基于误差平方和最小组合预测模型的各误差指标, 因而基于平均离散度的组合预测模型是一种有效的组合预测方法.

另外三种单项预测方法预测值序列与实际观测值序列的平均离散度分别为$f_1 = 0.2858$, $f_2 = 0.4439$和$f_3 = 0.5914$, 最优组合预测值序列与实际观测值序列的平均离散度为$f = 0.1914$, 满足$f < \min(f_1, f_2, f_3)$, 故由定义2.6知实例基于平均离散度的最优组合预测模型为优性组合预测.

§5. 结 语

在支撑度定义的基础上, 构建了平均指数支撑度最优组合预测模型和与之等价的平均离散度最优组合预测模型, 理论上研究平均离散度最优组合预测模型的非劣性组合预测和优性组合预测存在性、冗余预测方法的存在性及判定定理等性质. 最后的实证分析也说明基于平均离散度的组合预测模型是一种合理和有效的组合预测方法.

参 考 文 献

[1] Bates, J.M. and Granger, C.W.J., Combination of forecasts, *Operations Research Quarterly*, **20(4)**(1969), 451–468.

[2] Granger, C.W.J., Combining forecasts — twenty years later, *Journal of Forecasting*, **8(3)**(1989), 167–173.

[3] Donaldson, R.G. and Kamstra, M., Forecast combining with neural networks, *Journal of Forecasting*, **15(1)**(1996), 49–61.

[4] 唐小我, 马永开, 曾勇, 杨桂元, 现代组合预测和组合投资决策方法及应用研究, 北京: 科学出版社, 2003.

[5] 唐小我, 组合预测误差信息矩阵研究, 电子科技大学学报, **21(4)**(1992), 448–454.

[6] 马永开, 杨桂元, 唐小我, 非负权重组合预测的冗余定理, 系统工程理论方法应用, **4(4)**(1995), 33–39.

[7] 陈华友, 侯定丕, 基于预测有效度的优性组合预测模型研究, 中国科学技术大学学报, **32(2)**(2002), 172–180.

[8] 王应明, 基于相关性的组合预测方法研究, 预测, **21(2)**(2002), 58–62.

[9] 陈华友, 盛昭瀚, 刘春林, 基于向量夹角余弦的组合预测模型的性质研究, 管理科学学报, **9(2)**(2006), 1–8.

[10] 权双燕, 基于等维递补的多变量灰色组合预测模型, 纯粹数学与应用数学, **25(1)**(2009), 203–208.

[11] Yager, R.R., The power average operator, *IEEE Transactions on Systems, Man and Cybernetics*, **31(6)**(2001), 724–731.

[12] 安徽统计局, 安徽统计年鉴, 北京: 中国统计出版社, 2000–2009.

Research on the Optimal Combination Forecasting Model and Its Properties Based on Exponential Supporting Degree

Yuan Hongjun

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

Chen Huayou

(*School of Mathematical Science, Anhui University, Hefei, 230601*)

Hu Lingyun

(*School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, 233030*)

Based on the definition of supporting degree, we put forward the definitions of average index-supporting degree and average dispersion, and we build an optimal combination forecasting model based on exponential supporting degree. We also give its equivalent average dispersion optimal combination forecasting model. We propose the conception of superior combination forecasting, we obtain the results, including the sufficient conditions for the existing non-inferior and superior combination forecasting, redundant forecasting and the redundant information determination. Finally, we give the example to show that the model is effective.

Keywords: Combination forecast, average exponential supporting degree, average dispersion, superior combination forecasting.

AMS Subject Classification: 90B50.