第 27 卷　第 1 期
1998 年 2 月

电子科技大学学报
Journal of UEST of China

Vol. 27　No. 1
Feb. 1998

两种组合预测优化模型的分析和比较*

马永开**　　唐小我
（安徽财贸学院基础部　蚌埠　233041）（电子科技大学管理学院　成都　610054）

【摘要】　对两种线性组合预测方法的优化模型进行分析和比较，旨在为实际应用中的组合预测方法的优化设计提供理论基础，其结论对组合预测理论和实践有着重要意义。

关　键　词　组合预测；　优化模型；　有效组合；　无效组合
中图分类号　F201；　F224.0

经济系统往往是一个复杂系统。复杂系统内各种因素之间及其同外部因素之间存在着十分复杂的交互作用和因果关系，对于经济系统的预测仅靠一个预测模型和一种预测方法已难于完成，必须从不同角度出发，建立各类不同的预测模型，形成多个预测结果，然后再对这些预测结果使用组合预测方法。建立线性组合预测模型的一个重要步骤是对组合权重进行优化，选择不同的精度优化指标可导出不同的组合预测优化模型。在预测实践中，我们怎样选择组合预测优化模型对组合权重进行优化呢？为此，本文对两种常见的组合预测优化模型进行分析、比较研究，以期为线性组合预测模型设计过程中的优化模型的选择提供理论基础。

1　两种常见的组合预测优化模型及其分析

设某预测问题的统计观测值有 N 期，y_t 为第 t 期的统计观测值，对该预测问题有 m 种预测模型，f_{it}为第 i 种预测模型第 t 期的拟合预测值，其预测绝对误差 $e_{it}= f_{it}- y_t$（$i=1, 2, \cdots, m$；$t=1, 2, \cdots, N$）。用 f_t 表示组合预测模型第 t 期的拟合预测值，则 $f_t=k_1f_1t+k_2f_2t+\cdots+k_mf_mt$，其中 k_i 为第 i 种预测模型的组合权重；$E_t=f_t-y_t$ 为组合预测模型的第 t 期绝对误差，则 $E_t=k_1e_1t+k_2e_2t+\cdots+k_me_mt$，$t=1, 2, \cdots, N$。

1.1　一阶组合预测优化模型

选用绝对误差绝对值和 $E=\sum\limits_{t=1}^{N}|E_t|$ 作为组合预测模型精度优化指标，可得到如下的一阶组合预测优化模型。

模型（A）　$\min E=\sum\limits_{t=1}^{N}|E_t|=\sum\limits_{t=1}^{N}\left|\sum\limits_{i=1}^{N}k_ie_{it}\right|$

$$\text{s.t}\quad R_m^TK=1\qquad K\geqslant 0$$

其中　R_m 为元素全为 1 的 m 维列向量，$K=(k_1, k_2, \cdots, k_m)^T$ 为组合权重向量。

模型（A）可化为线性规划问题求解[1]，设模型（A）的最优解为 $K_A=(k_{A1}, k_{A2}, \cdots, k_{Am})^T$，记相对应的线性组合预测模型为 f_A，即

$$f_A:f_{At}=k_{A1}f_{1t}+k_{A2}f_{2t}+\cdots+k_{Am}f_{mt}$$

*　1997 年 6 月 26 日收稿，1997 年 9 月 16 日修改定稿
*　国家自然科学基金资助项目，自然科学基金号：79670013
**　男　34 岁　硕士　讲师

容易看出，模型(A)和下面的模型(A_1)等价。

模型(A_1)　$\min E=\frac{1}{N}\sum_{t=1}^{N}|E_t|$

　　　　　s.t　$R_m^T K=1$　　$K\geqslant 0$

由统计学知识可知，我们可以用 $E=\frac{1}{N}\sum_{t=1}^{N}|E_t|$ 作为线性组合预测模型未来一期预测精度的估计值(这里采用绝对误差绝对值作为预测模型的期精度指标)，模型(A)保证 f_A 是所有可能的线性组合中估计值最好的。

为了研究 f_A，这时引入下面的绝对误差矩阵 $\boldsymbol{A}$

$$\boldsymbol{A}=\begin{bmatrix} e_{11} & e_{12} & e_{13} & \cdots & e_{1N} \\ e_{21} & e_{22} & e_{23} & \cdots & e_{2N} \\ \cdots & \cdots\text{合预}\cdots & & \cdots & \cdots \\ e_{m1} & e_{m2} & e_{m3} & \cdots & e_{mN} \end{bmatrix}_{m\times N}$$

绝对误差矩阵 $\boldsymbol{A}$ 的第 i 行是第 i 种预测模型的各期绝对误差组成的绝对误差序列，$\boldsymbol{A}$ 的第 j 列分别是参加组合的 m 种预测模型在第 j 期上的绝对误差。

定理1　设绝对误差矩阵 $\boldsymbol{A}$ 满足：对任意 $j\in\{1,2,\cdots,N\}$，$\{e_{ij},e_{2j},\cdots,e_{mj}\}$ 为同号序列，且 $\sum_{t=1}^{N}|e_{mt}|<\sum_{t=1}^{N}|e_{it}|$ ($i=1,2,\cdots,m-1$)，则模型(A)有唯一最优解 $K_A=(0,0,\cdots,0,1)^T$，即 $f_{At}=f_{mt}$。

证明　由 $e_{1t},e_{2t},\cdots,e_{mt}$ 为同号序列知

$$|E_t|=|\sum_{i=1}^{k}k_i e_{it}|=\sum_{i=1}^{m}k_i|e_{it}|$$

对 $t=1,2,\cdots,N$ 都成立。所以

$$E=\sum_{t=1}^{N}|E_t|=\sum_{t=1}^{N}(\sum_{i=1}^{m}k_i|e_{it}|)=\sum_{i=1}^{m}k_i(\sum_{t=1}^{N}|e_{it}|)\geqslant$$
$$\sum_{i=1}^{m}k_i(\sum_{t=1}^{N}e_{mt})=(\sum_{i=1}^{m}k_i)(\sum_{t=1}^{N}e_{mt})=\sum_{t=1}^{N}e_{mt}$$

式中，等号仅在 $k_1=k_2=\cdots=k_{m-1}=0$，$k_m=1$ 时才成立。故当定理的条件成立时，模型(A)有唯一最优解 $K_A=(0,0,\cdots,0,1)^T$，此时 $f_{At}=f_{mt}$。

当组合预测模型的组合结构中只有一种预测模型时，我们就称该组合预测模型为无效组合，否则称为有效组合。定理1表明当参加组合的各种不同预测模型在任意确定的一期上的绝对误差符号相同，且在参加组合的各单项预测模型中绝对误差绝对值和最小的预测模型是唯一时，f_A 是无效组合。我们的研究表明，使 f_A 为无效组合的原因比较复杂，定理1给出了一种原因。

1.2　二阶组合预测优化模型

我们取绝对误差平方和作为组合预测模型精度优化指标，可得下面的二阶组合预测优化模型。

模型(B)　$\min J=\sum_{t=1}^{N}(E_t)^2=K^T B_m K$

　　　　s.t　$K^T R_m=1$　$K\geqslant 0$

其中　$B_m=\boldsymbol{A}\boldsymbol{A}^T=(b_{ij})_{m\times m}$ 是预测误差协方差阵，$b_{ij}=\sum_{t=1}^{N}e_{it}\cdot e_{jt}$ ($i,j=1,2,\cdots,m$)。

模型(B)是二次规划问题，可以通过 Kuhn-Tucker 条件化为线性规划问题求解[2]。设模型(B)的最优解为 K_B，以 K_B 作为组合权重的线性组合预测模型为 f_B。我们对模型(B)稍加改造，可

得与模型(B)等价的模型(B_1)如下：

模型(B_1)　$\min J=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(E_t)^2}$

$\text{s.t}\quad K^T R_m=1\quad K\geqslant 0$

因为参加组合的各单项预测模型在组合前都要经过无偏性调整, 所以, 我们认为 $f_{1t}, f_{2t}, \cdots, f_{mt}$都是无偏的, 此时有$\sum_{t=1}^{N}E_t=\sum_{t=1}^{N}(\sum_{i=1}^{m}k_ie_{it})=\sum_{i=1}^{m}k_i(\sum_{t=1}^{N}e_{it})=0, J=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(E_t)^2}=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(E_t-\frac{1}{N}E_t)^2}$就是线性组合预测模型 f_t 的各期绝对误差数据 $E_1, E_2, \cdots, E_N$ 的标准差, 它反映了 $E_1, E_2, \cdots, E_N$ 的离散状况, 也就反应了 f_t 各期绝对误差取值的稳定性, 模型(B)保证 f_B 的这种稳定性最好。

模型(B)是目前研究最多的组合预测优化模型, 有关它的理论是比较丰富的。这里, 我们给出 f_B 的一个性质。

定理 **2**　设 B_m 满足：$b_{mm}=\min\{b_{11}, b_{22}, \cdots, b_{mm}\}$, 则 f_B 为无效组合的充要条件是：$b_{mj}\geqslant b_{mm}, j=1, 2, \cdots, m$。

定理 2 的证明参见文献[3]。

B_m 中的元素 b_{ij}表示第 i 种模型和第 j 种预测模型的预测误差协方差($i\neq j$)。定理 2 表明当参加组合的所有单项预测模型和预测误差平方和最小的预测模型强正相关时, 它们和预测误差平方和最小的预测模型反映了预测对象的同一侧面, 所以 f_B 是无效组合。

2　两种组合预测优化模型的比较

模型(A)和模型(B)是常见的两种组合预测优化模型。前面, 我们已经分别对这两种模型从优化目标和无效组合的产生条件两个方面进行分析。下面, 我们从两个模型 的求解难度、对异常数据的灵敏度、对应的组合预测模型的稳定性等几个方面对这两个模型进行比较研究。

2.1　模型求解难度比较

模型(A)和模型(B)都可化为线性规划问题求解, 而求解线性规划问题有现成的计算机软件, 所以, 模型(A)和模型(B)的求解难度相当。

当模型(A)或模型(B)的最优解不唯一时, 就要进行二次决策, 从多个最优解中选择一个最适合预测对象的组合权重向量。文献[4]证明了当 B_m 正定时, 模型(B)有唯一最优解, 所以模型(A)较模型(B)更易出现多解情形。

2.2　对异常数据的敏感性比较

由于模型(B)的目标函数将组合预测模型的每一期绝对误差先平方再求和, 这种处理方法夸大了各期绝对误差大小的差距。从前面分析知, 模型(B)的优化目标是使组合预测模型的各期绝对误差取值的稳定性最好; 只要某单项预测模型在某一期上出现异常数据, 模型(B)为了消除异常数据的影响, 在其最优解 K_B 中给该预测模型以较小的权重, 甚至行使否决权。模型(A)的目标函数是组合预测模型的各期绝对误差绝对值和, 它如实地反映了各期绝对误差大小的差距。所以, 能比较客观地对待异常数据出现。当某种单项预测模型在某一期出现异常数据时, 模型(A)的最优解 K_A 会客观地作反应, 相对于模型(B)的最优解的夸大反应来说, 模型(A)的反应相对稳定。

2.3　相应的组合预测模型的精度稳定性比较

组合预测模型的精度信息包含在以下的期精度序列中

$$|E_1|, \quad |E_2|, \cdots, \quad |E_N| \tag{1}$$

将各期绝对误差加绝对值是因为预测模型的精度分析只考虑绝对误差大小，而不考虑绝对误差的符号。

依据统计学观点，精度序列(1)的均值 E 和标准差 S 能比较好地刻划精度序列(1)的特征，所以 E 和 S 能比较好地反应组合预测模型的精度特征。通常，我们可以用 E 反应组合预测模型未来一期精度的大小；用 S 反应组合预测模型的精度稳定性。这里，S 的表达式如下

$$S = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(|E_t| - E)^2} \tag{2}$$

从前面分析知，模型(B)实质上是将组合预测模型的各期绝对误差取值的稳定性作为优化的目标。而(2)不考虑绝对误差的正负号，仅考虑绝对误差绝对值。用 S 作为组合预测模型的稳定性指标，f_A 和 f_B 的稳定性状况会怎样呢？

定理 3　f_B 的精度稳定性优于 f_A 的精度稳定性。

证明
$$S = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(|E_t| - E)^2} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}E_t^2 - (E)^2} = \sqrt{\frac{1}{N}J - \frac{E^2}{N^2}}$$

设 f_A 的绝对误差值绝对值和为 E_A，绝对误差平方和为 J_A，稳定性指标为 S_A。f_B 的绝对误差绝对值和为 E_B，绝对误差平方和为 J_B，稳定性指标为 S_B。由模型(A)和模型(B)的意义知：$J_A \geqslant J_B$，$E_A \leqslant E_B$。所以

$$\frac{1}{N}J_A - \frac{E_A^2}{N^2} \geqslant \frac{1}{N}J_B - \frac{E_B^2}{N^2} \geqslant 0$$

$$S_A = \sqrt{\frac{1}{N}J_A - \frac{E_A^2}{N^2}} \geqslant \sqrt{\frac{1}{N}J_B - \frac{E_B^2}{N^2}} = S_B$$

S 的值越小，反应精度序列(1)中各期精度数据的起伏变化幅度越小，组合预测模型的精度稳定性越好，故定理 3 成立。

实际上，从模型(A_1)可知，模型(A)仅对 E 进行优化，所以忽略了 S 的优化，就造成 f_A 的稳定性差。而 $J = N((S)^2 + (E)^2)$，所以模型(B)和下面的模型(B_2)等价。

模型(B_2)　$\min J_2 = (S)^2 + (E)^2$

$\text{s.t.}\quad K^T R_m = 1 \quad K \geqslant 0$

从此模型可看出，模型(B)在对组合预测模型优化时，兼顾了 E 和 S，因此 f_B 的稳定性优于 f_A。

3 实　例

将表 1 中的 f_{1t} 和 f_{2t} 进行组合：$f_t = k_1 f_{1t} + k_2 f_{2t}$，使用模型($A$)和模型($B$)求得两个最优组合权重向量分别为：$K_A = (0, 1)^T$，$K_B = (0.611\,4, 0.388\,6)^T$，得到两种最优组合预测模型如下

$$f_A : f_{At} = f_{2t}$$

$$f_B : f_{Bt} = 0.611\,4 f_{1t} + 0.388\,6 f_{2t}$$

表 **1**　某省 **1980** 年～**1992** 年化工行业专门人才数的统计数据
及采用两种预测模型的拟合数据

时间/年	f_{1t}	f_{2t}	y_t	t
1980	5 266.75	4 348.80	6 014	1
1981	6 027.72	5 393.98	6 398	2
1982	6 888.12	6 520.19	6 181	3
1983	7 851.56	7 727.4	7 416	4
1984	8 915.62	9 015.65	9 076	5
1985	10 080.90	10 384.90	8 399	6
1986	11 347.40	11 835.20	9 215	7
1987	12 715.20	13 366.50	11 125	8
1988	14 184.27	14 978.80	14 120	9
1989	15 754.00	16 672.20	17 238	10
1990	17 425.90	18 446.50	18 689	11
1991	19 198.50	20 301.91	20 592	12
1992	21 072.40	22 238.30	22 665	13

本例的组合预测问题的绝对误差矩阵 A 满足定理 1 的条件，所以 f_A 是无效组合。由预测误差协方差阵 B_2 可看出，f_{1t}和f_{2t}虽然是正相关的，但不是强正相关的，所以 f_B 是有效组合。我们进一步计算 f_A 和 f_B 的 E 指标值和 S 指标值，并建立表 2，从表 2 可看出 f_B 要优于 f_A，因为 f_B 的 E 值仅比 f_A 增加 47.235 0，但稳定性指标 S 却下降了 208.575 7。此例表明，模型(B)往往比模型(A)有效。

表 **2**　两种最优组合预测模型的精度指标

	E	S
f_A	970.13	831.030 3
f_B	1017.455	622.454 6

参　考　文　献

1　徐大江. 确定最优组合预测权重数的线性规划方法. 预测, 1993, 12(2):56～57

2　曾　勇, 唐小我. 非负权重最优组合预测的计算方法研究. 统计研究, 1994, (3):70～74

3　马永开, 唐小我, 杨桂元. 优性组合预测方法存在判别定理, 预测, 1995, 14(2):57～58

4　唐小我, 曹长修, 金德运. 组合预测最优加权向量的进一步研究. 预测, 1994, 13(2):48～49

Analysis and Comparison Between Two Models of Optimizing Linear Combination Forecasting Model

Ma Yongkai
(Anhui Institute of Finance and Trade Bengbu 233041)
Tang Xiaowo
(Magagment College, UEST of China Chengdu 610054)

Abstract In this paper, two useful models that optimizt linear combination forecasting model are analysed and conpared in order to provide theory for a designer of combination forecasting model in practical application. It is meanimgful to both theory and practice of combinatel forecasting.

Key words combination forecasting; optimizing model; effective combination; invalid combination

编辑 徐安玉

·科研成果介绍·

磁光盘格式化仪

主研人员:高正平 周建平 王志刚 王 华 张 勋 罗 勇 黄晓革 张 鹰 陈宏猷等

磁光盘格式化仪利用磁光盘驱动器对磁光盘作低级格式化,并取出格式化期间的有关数据来分析和统计磁光盘的质量参数。系统涉及计算机技术、磁光盘驱动器技术、SCSI接口技术、磁光盘格式及信息处理技术、高速DSP技术。在模块结构上,采用主控计算机、扩展控制器、DSP电路、磁光驱动器四位一体的功能模块设计,易于扩展和升级,技术先进,性能稳定。

磁光盘格式化仪的研制成功,提供了1300 MB磁光盘产业所需的格式化及质量检测设备,节约了大量外汇,为研制、开发和生产更大容量系列磁光盘提供了技术保证。

该研制成果在磁光记录领域中磁光的低级格式化和质量检测设备方面填补了国内空白,达到国际90年代初先进水平。

·科 卞·