基于相关性的组合预测方法研究

王应明
(厦门大学 管理学院, 福建 厦门 361005)

摘 要: 本文对组合预测方法进行研究, 提出了四种基于相关性的组合预测方法, 即: 关联度极大化组合预测方法, 相关系数极大化组合预测方法, 夹角余弦极大化组合预测方法, Theil 不等系数极小化组合预测方法。四种基于相关性的组合预测方法均能够取得比较好的组合预测效果。

关键词: 组合预测; 灰关联度; 相关系数; 夹角余弦; Theil 不等系数
中图分类号: O221.1 文献标识码: A 文章编号: 1003-5192-(2002)02-0058-05

Research on the Methods of Combining Forecasts Based on Correlativity

WANG Ying-ming
(*School of Management, Xiamen University, Xiamen* 361005, *China*)

Abstract: This paper studies the methodology of combining forecasts. Four new methods are proposed, which are maximum grey correlation degree method, maximum correlation coefficient and included angle cosine methods as well as minimum Theil coefficient method. All the methods can lead to satisfactory forecasting results.

Key words: combining forecasts; grey correlation degree; correlation coefficient; included angle cosine; Theil coefficient

1 引言

组合预测由于比单方法预测更有效、能提高模型的拟合精度和预测能力, 因此, 自从 1969 年问世以来, 一直是国内外预测界研究的热点课题, 并在世界各国范围内得到广泛应用。综观国内外研究现状可以发现, 现有的组合预测方法都是基于直接改善某种拟合误差角度提出来的。笔者认为, 这不是研究组合预测方法的唯一途径、也未必就是最佳途径。另外, 根据预测惯例, 预测效果的评价至少可以从平方和误差(*SSE*)、平均绝对误差(*MAE*)、均方误差(*MSE*)、平均绝对百分比误差(*MAPE*)、均方百分比误差(*MSPE*)等五个方面进行全方位的综合衡量, 任何一种仅以其中某个误差指标为最优进行组合预测的方法都不是很全面的, 因此, 有必要研究新的组合预测方法以便能使各种误差指标皆得到不同程度的的改善。本文拟从相关性指标(如关联度、相关系数、夹角余弦、Theil 不等系数等)的角度进行组合预测方法的研究。新的预测方法不直接考虑预测误差的大小, 与传统的组合预测方法有较大的差别。

2 相关性组合预测原理及参数估计

设某预测问题在某一时段的实际值为 y_t ($t=1, 2, \cdots, n$), 对此预测问题有 m 种可行的预测方法, 其预测值或模型拟合值分别为 f_{tj} ($t=1, 2, \cdots, n$; $j=1, 2, \cdots, m$)。又设 m 种预测方法的组合权向量为 $W=(W_1, W_2, \cdots, W_m)^T$, 且满足归一化约束条件和非负约束条件

$$e^T W=1 \tag{1}$$

$$W\geq 0 \tag{2}$$

其中 $e^T=(1, 1, \cdots, 1)$。根据组合预测原理, 我们可以有三种不同的组合方式, 即: 算术平均组合、几何平均组合、调和平均组合, 其组合公式分别为

$$\hat{y_t}=\frac{\sum_{j=1}^{m} W_j f_{tj}}{\sum_{j=1}^{m} W_j}=\sum_{j=1}^{m} W_j f_{tj} \qquad t=1, 2, \cdots, n \tag{3}$$

$$\hat{y_t}=\sqrt[\sum_{j=1}^{m} W_j]{\prod_{j=1}^{m} f_{tj}^{W_j}}=\prod_{j=1}^{m} f_{tj}^{W_j} \qquad t=1, 2, \cdots, n \tag{4}$$

收稿日期: 2000-11-18
基金项目: 霍英东教育基金会资助项目(71080)

$$\hat{y_t}=\frac{\sum_{j=1}^{m}W_j}{\sum_{j=1}^{m}\frac{W_j}{f_{tj}}}=\frac{1}{\sum_{j=1}^{m}\frac{W_j}{f_{tj}}}\qquad t=1,2,\cdots,n \tag{5}$$

其中 $\hat{y}$ 为 t 时刻的组合预测值。很显然,不管采用何种组合方式、也不论采取何种非负加权系数,下列不等式始终成立

$$\min_j f_{tj}\leqslant\hat{y_t}\leqslant\max_j f_{tj}\qquad t=1,2,\cdots,n \tag{6}$$

根据灰关联度原理,若组合预测模型与实际数据完全吻合(即100%准确),则模型预测值与实际值两个序列$\{\hat{y_t}\}\{y_t\}$之间的灰关联度必为1(即达到最大值),也就是说,模型越准确、灰关联度越大,因此,我们可以令组合预测模型与原始数据两个序列之间的灰关联度最大,以此为基础来建立相关性组合预测方法。类似的组合预测方法还有"相关系数极大化组合预测方法"、"夹角余弦极大化组合预测方法"和"Theil不等系数极小化组合预测方法"等,此处一并介绍。为了介绍方便起见,我们引进两个集合符号:$\Omega_1=\{1,2,\cdots,m\}$,$\Omega_2=\{1,2,\cdots,n\}$;同时令

$$\gamma_{0j}=\frac{1}{n}\sum_{t=1}^{n}\frac{\min_{j\in\Omega_1}\min_{t\in\Omega_2}|y_t-f_{tj}|+\rho\max_{j\in\Omega_1}\max_{t\in\Omega_2}|y_t-f_{tj}|}{|y_t-f_{tj}|+\rho\max_{j\in\Omega_1}\max_{t\in\Omega_2}|y_t-f_{tj}|}\qquad j=1,2,\cdots,m \tag{7}$$

其中 γ_{0j}为单个预测方法预测值f_{tj}与实际值y_t两个序列之间的关联度,$0<\gamma_{0j}<1$;$\rho\in(0,1)$为分辨率系数,通常可取0.5。由于不等式(6)的存在,组合预测值$\hat{y_t}$与实际值y_t两个序列之间的关联度为

$$\gamma=\frac{1}{n}\sum_{t=1}^{n}\frac{\min_{j\in\Omega_1}\min_{t\in\Omega_2}|y_t-f_{tj}|+\rho\max_{j\in\Omega_1}\max_{t\in\Omega_2}|y_t-f_{tj}|}{|y_t-\hat{y_t}|+\rho\max_{j\in\Omega_1}\max_{t\in\Omega_2}|y_t-f_{tj}|} \tag{8}$$

很显然,灰关联度 γ 是组合权向量 W 的函数,为了得到最佳的组合预测效果,灰关联度 γ 应该愈大愈好。为此,构造如下最优化模型

$$\max\gamma=\sum_{t=1}^{n}\frac{\min_{j\in\Omega_1}\min_{t\in\Omega_2}|y_t-f_{tj}|+\rho\max_{j\in\Omega_1}\max_{t\in\Omega_2}|y_t-f_{tj}|}{|y_t-\hat{y_t}|+\rho\max_{j\in\Omega_1}\max_{t\in\Omega_2}|y_t-f_{tj}|} \tag{9}$$

$$\text{s.t.}\begin{cases}e^{T}W=1 & (10)\\ W\geqslant 0 & (11)\end{cases}$$

该模型为最简单的线性约束非线性规划模型,采用简约梯度方法求解甚是方便,也可借助于专门求解非线性规划的数学软件 Lingo 求解;式中 $\hat{y_t}$ 可取(3)、(4)、(5)式中的任何一种组合形式代入。

当从相关分析角度考察组合预测问题的时候,我们自然而然是希望组合预测值$\hat{y_t}$与实际值y_t两个序列之间是高度线性相关,相关系数愈大,表明组合预测效果愈佳;当它们在整个时段内完全吻合时,相关系数达到理想的最大值1,但这实际上是不可能的。为此,我们令

$$R_j=\frac{\sum_{t=1}^{n}(y_t-\bar{y})\cdot(f_{tj}-\bar{f_j})}{\sqrt{\sum_{t=1}^{n}(y_t-\bar{y})^2}\cdot\sqrt{\sum_{t=1}^{n}(f_{tj}-\bar{f_j})^2}}\qquad j=1,2,\cdots,m \tag{12}$$

$$\hat{R}=\frac{\sum_{t=1}^{n}(y_t-\bar{y})\cdot(\hat{y_t}-\bar{\hat{y}})}{\sqrt{\sum_{t=1}^{n}(y_t-\bar{y})^2}\cdot\sqrt{\sum_{t=1}^{n}(\hat{y_t}-\bar{\hat{y}})^2}} \tag{13}$$

其中 R_j 为单个预测方法预测值f_{tj}与实际值y_t两个序列之间的相关系数,$-1<R_j<1$;$\hat{R}$ 为组合预测值$\hat{y_t}$与实际值y_t两个序列之间的相关系数;$\bar{y}$、$\bar{f_j}$和$\bar{\hat{y}}$分别为y_t、f_{tj}和$\hat{y_t}$的平均值,它们的算法应该与$\hat{y_t}$的组合形式保持一致。也就是说,如果$\hat{y_t}$采用算术平均组合形式,则$\bar{y}$、$\bar{f_j}$、$\bar{\hat{y}}$采用算术平均值;如果$\hat{y_t}$采用几何平均组合形式,则$\bar{y}$、$\bar{f_j}$、$\bar{\hat{y}}$采用几何平均值;如果$\hat{y_t}$采用调和平均组合形式,则$\bar{y}$、$\bar{f_j}$、$\bar{\hat{y}}$也必须采用调和平均值。为了得到最佳的组合预测效果,我们求解如下最优化模型可得到最优组合权向量 W^*

$$\max\hat{R}=\frac{\sum_{t=1}^{n}(y_t-\bar{y})\cdot(\hat{y_t}-\bar{\hat{y}})}{\sqrt{\sum_{t=1}^{n}(y_t-\bar{y})^2}\cdot\sqrt{\sum_{t=1}^{n}(\hat{y_t}-\bar{\hat{y}})^2}} \tag{14}$$

$$\text{s.t.}\begin{cases}e^{T}W=1 & (15)\\ W\geqslant 0 & (16)\end{cases}$$

另外,我们还可以从夹角余弦的角度来考虑组合预测问题。如果我们将实际值$y_t(t=1,2,\cdots,n)$和组合预测值$\hat{y_t}(t=1,2,\cdots,n)$分别看成是两个向量,即:$Y=(y_1,y_2,\cdots,y_n)^T$和$\hat{Y}=(\hat{y_1},\hat{y_2},\cdots,\hat{y_n})^T$,在不存在预测误差的情况下,这两个向量应该方向一致,其夹角为零;但事实上,预测误差的存在是不可避免的,在这种情况下,我们自然而然是希望Y和$\hat{Y}$两个向量之间的夹角愈小愈好。用数学的语言来描述,也就是夹角的余弦愈大愈

好。为此,我们令

$$\eta_j=\frac{\sum_{t=1}^{n} y_t \cdot f_{tj}}{\sqrt{\sum_{t=1}^{n} y_t^2} \cdot \sqrt{\sum_{t=1}^{n} f_{tj}^2}} \qquad j=1,2,\cdots,m \tag{17}$$

$$\eta=\frac{\sum_{t=1}^{n} y_t \cdot \hat{y_t}}{\sqrt{\sum_{t=1}^{n} y_t^2} \cdot \sqrt{\sum_{t=1}^{n} \hat{y}_t^2}} \tag{18}$$

其中 η_j 为单个预测方法预测值 f_{tj} 与实际值 y_t 两个序列之间的夹角余弦, $0<\eta_j<1$; η 为组合预测值 $\hat{y_t}$ 与实际值 y_t 两个序列之间的夹角余弦。为了得到最佳的组合预测效果,求解如下最优化模型可得到最优组合权向量 W^*

$$\max\ \eta=\frac{\sum_{t=1}^{n} y_t \cdot \hat{y_t}}{\sqrt{\sum_{t=1}^{n} y_t^2} \cdot \sqrt{\sum_{t=1}^{n} \hat{y}_t^2}} \tag{19}$$

$$\text{s.t.}\begin{cases} e^{\mathrm{T}} W=1 & (20)\\ W\geqslant 0 & (21)\end{cases}$$

另外, Theil 不等系数常常可以用来评价预测效果的好坏,计算公式为

$$\theta_j=\frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-f_{tj})^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t^2}+\sqrt{\frac{1}{n}\sum_{t=1}^{n} f_{tj}^2}} \qquad j=1,2\cdots,m \tag{22}$$

$$\theta=\frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y_t})^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t^2}+\sqrt{\frac{1}{n}\sum_{t=1}^{n} \hat{y}_t^2}} \tag{23}$$

其中 θ_j 为单个预测方法预测值 f_{tj} 与实际值 y_t 两个序列之间的 Theil 不等系数, $0<\theta_j<1$; θ 为组合预测值 $\hat{y_t}$ 与实际值 y_t 两个序列之间的 Theil 不等系数。很显然, Theil 不等系数总是愈小愈好。因此,为了得到最佳的组合预测效果,我们构造如下最优化模型

$$\min\theta=\frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y_t})^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n} y_t^2}+\sqrt{\frac{1}{n}\sum_{t=1}^{n} \hat{y}_t^2}} \tag{24}$$

$$\text{s.t.}\begin{cases} e^{\mathrm{T}} W=1 & (25)\\ W\geqslant 0 & (26)\end{cases}$$

解此最优化模型,可以很方便地得到最优组合权向量 W^*。

3　相关性组合预测效果评价

基于相关性的组合预测方法虽然不涉及模型拟合误差的好坏,但是,为了便于检验和评价相关性组合预测效果的好坏,本文仍旧按照预测效果评价原则和惯例,采用下列 5 项拟合误差指标作为评判准则,对预测效果进行全方位的综合性评价。

(1)平方和误差

$$SSE=\sum_{t=1}^{n}(y_t-\hat{y_t})^2 \tag{27}$$

其中 y_t 为预测事物实际观测值, $\hat{y_t}$ 为预测值。

(2)平均绝对误差

$$MAE=\frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y_t}| \tag{28}$$

(3)均方误差

$$MSE=\frac{1}{n}\sqrt{\sum_{t=1}^{n}(y_t-\hat{y_t})^2} \tag{29}$$

(4)平均绝对百分比误差

$$MAPE=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y_t}}{y_t}\right| \tag{30}$$

(5)均方百分比误差

$$MSPE=\frac{1}{n}\sqrt{\sum_{t=1}^{n}\left(\frac{y_t-\hat{y_t}}{y_t}\right)^2} \tag{31}$$

4　实证研究

已知某两个预测问题原始数据如表 1 所示,根据表 1 数据计算的各单个预测方法预测值与实际值之间的灰关联度、相关系数、夹角余弦、Theil 不等系统如表 2 所示,基于各种不同组合形式的相关性组合预测效果如表 3 所示,两表中各种灰关联度的计算均是在分辨率系数取 $\rho=0.5$ 时的计算结果。不过,实际研究中我们对分辨率系数 ρ 从 0.1 到 1.0 取了 10 组数据分别进行了测算,发现组合权向量 W 对分辨率系数 ρ 的变化极不敏感,在整个 $\rho\in(0.1,1)$ 的区间内,最佳组合权向量 W^* 几乎不变,这也许就是灰色系统理论推荐使用 $\rho=0.5$ 的缘故。另外,从表 3 的预测评价效果可以看出,基于相关性的组合预测效果是令人满意的,这说明运用相关性组合预测方法进行组合预测,从理论上和实践上都是行得通的。

表 **1**　预测实例原始数据

	t	1	2	3	4	5	6	7	8	9	10	11	12
例 1	y_t	3306	4318	4228	4846	6054	8084	8910	10247				
	f_{t1}	2988	4303	4767	5161	6289	7294	9371	9892				
	f_{t2}	3331	3991	4342	5886	5818	8414	8814	9564				
例 2	y_t	11.49	13.06	15.34	20.58	23.28	26.46	27.33	34.22	40.19	53.37	77.79	100.63
	f_{t1}	18.47	14.54	12.84	13.28	16.15	21.16	28.40	37.87	49.58	63.53	79.00	98.12
	f_{t2}	10.03	11.23	15.24	18.67	27.78	26.36	29.67	27.40	42.73	47.36	71.00	109.32

表 **2**　单方法预测相关指标计算

方法 / 指标	例 1		例 2	
	方法(Ⅰ)	方法(Ⅱ)	方法(Ⅰ)	(Ⅱ)方法
关联度	0.6323	0.6813	0.5739	0.6597
相关系数	0.9832	0.9801	0.9783	0.9870
夹角余数	0.9979	0.9974	0.9925	0.9951
Theil	0.0325	0.0360	0.0628	0.0497

表 **3**　组合预测效果评价

预测效果评价指标				*SSE*	*MAE*	*MSE*	*MAPE*	*MSPE*
例 1	个体预测		方法(Ⅰ)	1508966	378.50	153.55	0.0644	0.0264
			方法(Ⅱ)	1842451	356.38	169.67	0.0603	0.0307
	关联度极大化组合预测	算术平均组合 (γ= 0.7496)	W_1= 0.2947 W_2= 0.7053	1159074	266.06	134.58	0.0482	0.0248
		几何平均组合 (γ= 0.7503)	W_1= 0.2658 W_2= 0.7342	1166550	266.22	135.01	0.0482	0.0248
		调和平均组合 (γ= 0.7512)	W_1= 0.2658 W_2= 0.7342	1174594	266.37	135.47	0.0482	0.0248
	相关系数极大化组合预测	算术平均组合 (R= 0.9902)	W_1= 0.5400 W_2= 0.4600	963044	289.32	122.67	0.0512	0.0225
		几何平均组合 (R= 0.9906)	W_1= 0.5312 W_2= 0.4688	958479	288.01	122.38	0.0509	0.0223
		调和平均组合 (R= 0.9914)	W_1= 0.5221 W_2= 0.4779	954772	286.63	122.14	0.0506	0.0222
	夹角余弦极大化组合预测	算术平均组合 (η= 0.9987)	W_1= 0.5625 W_2= 0.4375	962034	293.68	122.60	0.0518	0.0224
		几何平均组合 (η= 0.9987)	W_1= 0.5449 W_2= 0.4551	957533	290.67	122.32	0.0513	0.0223
		调和平均组合 (η= 0.9987)	W_1= 0.5273 W_2= 0.4727	954268	287.65	122.11	0.0508	0.0222
	Theil 不等系数极小化组合预测	算术平均组合 (ϑ= 0.0260)	W_1= 0.5581 W_2= 0.4419	962007	292.83	122.60	0.0517	0.0224
		几何平均组合 (ϑ= 0.0260)	W_1= 0.5499 W_2= 0.4501	957451	291.63	122.31	0.0515	0.0223
		调和平均组合 (ϑ= 0.0259)	W_1= 0.5431 W_2= 0.4569	953692	290.69	122.07	0.0512	0.0221

续表3

预测效果评价指标				*SSE*	*MAE*	*MSE*	*MAPE*	*MSPE*
例2	个体预测		方法(Ⅰ)	401.56	4.8817	1.6699	0.1959	0.0731
			方法(Ⅱ)	245.58	3.5908	1.3059	0.0998	0.0334
	关联度极大化组合预测	算术平均组合(γ=0.7389)	W_1=0.3717 W_2=0.6283	96.34	2.2984	0.8179	0.0753	0.0264
		几何平均组合(γ=0.7282)	W_1=0.4067 W_2=0.5933	96.62	2.3601	0.8191	0.0788	0.0273
		调和平均组合(γ=0.7176)	W_1=0.4424 W_2=0.5576	101.40	2.4572	0.8392	0.0840	0.0290
	相关系数极大化组合预测	算术平均组合(R=0.9949)	W_1=0.4095 W_2=0.5905	94.90	2.3373	0.8118	0.0788	0.0282
		几何平均组合(R=0.9953)	W_1=0.4124 W_2=0.5876	96.52	2.3704	0.8717	0.0795	0.0275
		调和平均组合(R=0.9958)	W_1=0.4169 W_2=0.5831	101.32	2.4785	0.8388	0.0822	0.0280
	夹角余弦极大化组合预测	算术平均组合(η=0.9982)	W_1=0.4189 W_2=0.5811	94.93	2.3545	0.8119	0.0800	0.0286
		几何平均组合(η=0.9981)	W_1=0.4248 W_2=0.5752	94.48	2.3927	0.8185	0.0811	0.0280
		调和平均组合(η=0.9980)	W_1=0.4338 W_2=0.5662	101.26	2.4646	0.8386	0.0834	0.0287
	Theil不等系数极小化组合预测	算术平均组合(θ=0.0308)	W_1=0.4133 W_2=0.5867	94.89	2.3442	0.8118	0.0792	0.0284
		几何平均组合(θ=0.0311)	W_1=0.4218 W_2=0.5782	96.46	2.3873	0.8185	0.0807	0.0279
		调和平均组合(θ=0.0319)	W_1=0.4293 W_2=0.5707	101.23	2.4683	0.8385	0.0831	0.0285

参　考　文　献:

[1] 唐小我.经济预测与决策新方法及其应用研究[M].成都:电子科技大学出版社,1997.

[2] 项静恬.非线性复杂系统的综合技术(Ⅰ)——复杂系统的模型综合与组合预测[J].数理统计与管理,1995,14(1):59-64.

[3] 王应明,傅国伟.基于不同误差准则和范数的组合预测方法研究[J].控制与决策,1994,9(1):20-28.

[4] 王应明.广义加权多重平均组合预测技术研究[J].预测,1997,16(1):46-48.

[5] 王应明,罗英.广义加权算术平均组合预测技术研究[J].预测,1998,17(1):51-53,67.

[6] 王应明,罗英.基于调和平均数的组合预测方法研究[J].统计研究,1997,(2):66-68.

(上接46页)

(3)β系数在越短的时期内越稳定,投资者进行短期投资时,可以用β系数作为决策的依据。

(4)利用股票组合可以提高β系数的稳定性,可见,投资组合的确可以分散风险。

参　考　文　献:

[1] 靳云汇,李学.中国股市β系数的实证检验[J].数量经济技术经济研究,2000,(1):18-25.

[2] Sharp W F. Capital asset prices: a theory of market equilibrium under risk[J]. Journal of Finance, 1964, 19: 425-442.

[3] Lintner J. The valuation of risky assets and the selection of risky investment in stock portfolio and capital budgets[J]. Review of Economics and Statistics, 1965, 47: 13-37.

[4] 陆懋祖.高等时间序列经济计量学[M].上海:上海人民出版社,1999.57-110.

[5] Dickey S. Testing for unit roots in antoregressive moving average models of unknown order[J]. Biometrika, 1984, 71: 599-607.