

简单平均组合预测有效性的应用分析*

曾勇**

李玉东

(电子科技大学管理学院 成都 610054) (海军指挥自动化研究所 北京 100036)

唐小我

(电子科技大学管理学院 成都 610054)

【摘要】 通过实例分析,说明了当序列模式变动较大时,简单平均组合预测模型相对于其他组合预测模型的优越性,并且基于样本段的拟合精度不足以说明组合预测模型的外推预测精度。文中的分析对于组合预测模型的选择和应用具有实际参考价值。

关键词 组合预测; 简单平均组合; 拟合精度; 预测精度; 模型不确定性

中图分类号 F201; F224

由 Bates 和 Granger 提出的通过综合各单项预测的信息,从而有效提高预测精度的组合预测思想已被国内外预测学界广泛接受并得到了大量的研究和应用。各种组合框架和具体方法也相继提出,包括线性与非线性组合、贝叶斯与非贝叶斯组合以及适应性与非适应性组合等。在众多的组合预测方法中,简单方法与复杂方法究竟孰优孰劣的问题一直没有定论,这与预测学界围绕简单模型与复杂模型的长期争论是一致的^[1-3]。复杂的模型往往能够很好地拟合样本数据,并通过显著性很强的假设检验,但由于将相同数据段同时用于参数估计和模型结构识别检验的传统做法,更由于模型不确定性的存在^[4],所得模型很可能仅具有预测精度高的假象,实际的预测精度可能要低得多。关于组合预测的早期实证检验已表明^[5,6],当序列模式变化明显时,简单的组合预测模型常常表现出良好的外推预测性能。

迄今的国内组合预测研究在模型构造、权重确定和模型比较方面还主要侧重于拟合精度,对外推预测精度的考察尚少^[7-10]。本文通过一个实际案例的分析,说明了当被预测变量趋势变动显著时,简单平均组合相对于其他组合模型在外推预测方面具有较大的优越性。

1 模型

假定对于被预测变量 y_t 有 n 个单项预测 f_{it} , $i=1, 2, \dots, n$ 。本文涉及的组合预测模型如下:

1) 简单平均组合^[5]

$$f_{ct}^{(1)} = \sum_{i=1}^n w_i^{(1)} f_{it} \quad (1)$$

式中 $w_i^{(1)} = \frac{1}{n}$, $1 \leq i \leq n$ 。

2) 最优加权平均组合^[7]

$$f_{ct}^{(2)} = \sum_{i=1}^n w_i^{(2)} f_{it} \quad (2)$$

式中 $\sum_{i=1}^n w_i^{(2)} = 1$; $w_n^{(2)} = (w_1^{(2)}, w_2^{(2)}, \dots, w_n^{(2)})^T = E_n^{-1} R_n / R_n^T E_n^{-1} R_n$; $R_n = (1, 1, \dots, 1)^T$; $E_n = (e_{ij})_{n \times n}$; $e_{ij} =$

$\frac{1}{N} \sum_{t=1}^N (y_t - f_{it})(y_t - f_{jt})$; N 为样本数据长度。

3) 考虑系统偏差的最优加权平均组合^[11]

$$f_{ct}^{(3)} = \sum_{i=1}^n w_i^{(3)} (\bar{e}_i + f_{it}) \quad (3)$$

1998年7月7日收稿,1998年11月2日修改定稿

* 国家杰出青年科学基金资助项目,基金号:79725002

** 男 35岁 硕士 教授

式中 $\sum_{i=1}^n w_i^{(3)} = 1$; $w_n^{(3)} = (w_1^{(3)}, w_2^{(3)}, \dots, w_n^{(3)})^T = V_n^{-1} R_n / R_n^T V_n^{-1} R_n$; $V_n = (v_{ij})_{n \times n}$; $v_{ij} = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{e}_i - f_{it}) \times (y_t - \bar{e}_j - f_{jt})$; \bar{e}_i 为 f_{it} 的样本预测误差均值。

4) 考虑系统偏差和尺度偏差的最优加权平均组合^[12]

$$f_{ci}^{(4)} = \sum_{i=1}^n w_i^{(4)} (\alpha_i + \beta_i f_{it}) \quad (4)$$

式中 $\sum_{i=1}^n w_i^{(4)} = 1$; $w_n^{(4)} = (w_1^{(4)}, w_2^{(4)}, \dots, w_n^{(4)})^T = \sum_n^{-1} R_n / R_n^T \sum_n^{-1} R_n$; $\sum_n = (\sigma_{ij})_{n \times n}$; $\sigma_{ij} = \frac{1}{N} \sum_{t=1}^N (y_t - \alpha_i - \beta_i f_{it}) \times (y_t - \alpha_j - \beta_j f_{jt})$; α_i 、 β_i 分别为 f_{it} 的样本位置偏差和尺度偏差系数; α_i 和 β_i 由回归方程 $y_t = \alpha_i + \beta_i f_{it} + \varepsilon_{it}$ 最小二乘估计而得。

5) 线性回归组合^[13]

$$f_{ci}^{(5)} = w_0^{(5)} + \sum_{i=1}^n w_i^{(5)} f_{it} \quad (5)$$

式中 $w_0^{(5)}, w_1^{(5)}, \dots, w_n^{(5)}$ 由最小二乘估计而得。模型(3)-(5)的理论分析参见文献[12]。

2 实例分析

2.1 数据

本文的数据来源于文献[11]中英国 1977 年第 1 季度至 1985 年第 2 季度的季度通货膨胀率, 五个单项预测分别由五家机构作出, 简记为 HCF、LBS、NI、OECD 和 PD, 数据如表 1 所示。

2.2 组合预测结果与分析

将上述的五个模型应用于表 1 数据, 可得各模型在不同样本段与预测段的精度如表 2 所示。

从表 2 可知, 简单平均组合在各预测段的预测精度均高于其他组合预测模型, 这种优越性在预测段增加(样本段减少)时体现得尤为显著。如当样本段长度为 14 时, 虽然模型(4)和(5)具有明显低于模型(1)的拟合精度, 但其预测精度却大大低于模型(1)。随着样本段长度的增加, 简单平均组合在外推预测方面的相对优势有所下降, 特别是模型(4)和(5)的预测精度相对于模型(1)有明显改善, 当样本段长度为 26 时, 模型(4)的预测精度已接近模型(1)。

从表 2 还可以看出, 对于相同的数据分段, 拟合精度与预测精度呈现出并不一致的特点, 拟合精度高的模型其预测精度大多相对较低, 虽然这一现象不具有普遍性, 但至少说明仅仅依据模型样本段的拟合精度或假设检验不足以说明其在外推预测时的性能, 特别是在数据趋势存在较大的变动时, 这一问题更为突出, 这可以从本例数据变动趋势中得到说明。

从表 1 可知, 实际通胀率经历了从 1977 年第 1 季度到 1978 年第 2 季度相对平稳的上升后, 从 1978 年第 3 季度开始快速攀升到 1979 年的最高点, 之后又快速回落到 1978 年第 2 季度的水平(1980 年 2 季度), 经过短暂徘徊和相对的下降后, 从 1982 年第 2 季度开始, 稳定的通胀率一直持续到整个数据段末。伴随实际数据的变动, 各单项预测也经历了起伏波动的过程, 并且各单项预测的性能表现出不平稳性, 如表 3 和表 4 看出。

从表 3 和表 4 可知, 各单项预测样本段与预测段的性能存在严重的差异, 这将造成对样本数据拟合程度高的模型预测性能反而不佳。

从不同数据段考察, 五个模型各自拟合精度与预测精度的变化关系具有不同的特点。随着样本段长度的增加, 虽然各模型预测精度均随之提高, 但模型(1)和(2)的拟合精度与预测精度同时提高, 而模型(4)和(5)的拟合精度与预测精度呈相反方向变化, 模型(3)介于其中。这一现象也可以从数据趋势的变动和模型的复杂程度加以解释(本文涉及的五个组合预测模型的复杂程度是依次增加的。线性回归组合等价于无偏性调整后的单项预测与被预测变量无条件期望值的最优加权平均组合^[2])。当校本数据量少时, 所包含的趋势变动也少, 对样本数据拟合能力强的复杂组合预测模型自然具有很高的拟合精度, 但其估计的组合权重对样本数据依赖程度也高, 难以适应预测段序列模式的变化。这一点模型(4)

和(5)在样本段长度为14时表现最充分。在1980年第2季度以前,通胀率一直呈上升趋势,模型(4)和(5)很好地拟合了这一趋势。但这之后,通胀率转呈下降趋势,模型(4)和(5)无法适应这种变化,预测性能很差。随着样本数据量的增加,所包含的趋势变动也大,复杂程度高的组合预测模型拟合精度也随之下降。相反,对样本数据依赖程度低的简单组合预测模型随着序列模式的稳定,其拟合精度逐步提高。在本例中,从1980年第3季度即第15个数据开始,通胀率平稳下降后趋于稳定。

表1 1997(I)~1985(II)英国通货膨胀率及预测

年/季度	HCF	LBS	NI	OCED	PD	实际值
1977/1	9.960 00	10.559 00	10.194 00	12.875 00	14.900 00	9.868 70
1977/2	10.900 00	13.243 00	9.435 00	11.500 00	14.700 00	8.618 62
1977/3	10.650 00	12.320 00	10.611 00	10.750 00	15.900 00	8.800 60
1977/4	11.625 00	15.164 00	8.016 00	8.500 00	12.600 00	9.597 65
1978/1	9.900 00	10.359 00	8.931 00	9.500 00	9.500 00	10.152 82
1978/2	8.940 00	9.728 00	12.002 00	10.500 00	7.300 00	11.106 27
1978/3	8.900 00	11.494 00	11.433 00	8.750 00	7.600 00	15.758 54
1978/4	8.900 00	9.363 00	9.318 00	7.500 00	8.100 00	16.832 79
1979/1	9.700 00	10.526 00	9.705 00	8.7500 00	8.600 00	17.903 84
1979/2	10.875 00	10.191 00	10.196 00	8.500 00	9.900 00	19.215 09
1979/3	14.387 50	12.575 00	10.608 00	12.750 00	15.800 00	14.562 46
1979/4	14.900 00	12.717 00	15.268 00	12.000 00	17.599 99	13.728 12
1980/1	17.700 00	15.642 00	13.492 00	13.750 00	18.599 99	12.025 79
1980/2	17.974 99	14.973 00	15.239 00	13.750 00	19.500 00	11.178 51
1980/3	16.149 99	14.796 00	14.389 60	16.124 98	16.899 99	11.256 77
1980/4	12.500 00	10.000 00	11.235 00	14.250 00	14.200 00	11.026 58
1981/1	11.400 00	9.000 00	8.170 00	12.000 00	12.400 00	10.830 70
1981/2	9.462 50	9.268 00	8.357 00	10.500 00	10.700 00	9.150 82
1981/3	11.175 00	8.571 00	10.132 00	10.250 00	10.500 00	7.974 28
1981/4	10.925 00	10.455 00	10.107 00	8.625 00	11.000 00	6.960 84
1982/1	11.300 00	8.929 00	9.771 40	9.750 00	12.300 00	6.001 91
1982/2	9.650 00	7.792 00	8.311 00	8.375 00	8.500 00	4.814 54
1982/3	9.050 00	7.328 00	7.854 00	8.125 00	8.500 00	4.777 31
1982/4	8.137 50	6.734 90	5.051 00	7.500 00	6.000 00	4.562 60
1983/1	6.387 50	6.327 00	7.836 00	5.375 00	5.900 00	4.621 82
1983/2	6.200 00	6.820 00	7.943 14	5.000 00	4.700 00	4.999 96
1983/3	6.062 50	6.125 80	7.370 60	6.250 00	5.200 00	4.719 43
1983/4	6.012 50	6.007 00	6.636 00	5.500 00	5.300 00	4.730 58
1984/1	5.600 00	5.873 00	7.115 50	5.750 00	5.400 00	5.230 11
1984/2	5.300 00	5.849 00	6.226 60	5.375 00	4.600 00	5.014 13
1984/3	5.287 50	4.908 90	6.250 00	5.250 00	5.530 00	5.352 18
1984/4	5.990 00	4.380 00	6.411 20	5.250 00	5.430 00	5.360 61
1985/1	5.790 00	5.200 00	7.201 00	5.250 00	5.390 00	4.761 35
1985/2	5.750 00	5.200 00	4.339 00	4.750 00	5.560 00	3.894 70

表2 组合预测模型样本段与预测段的根均方差

模型	数 据 段							
	1~14	15~34	1~18	19~34	1~22	23~34	1~26	27~34
(1)	4.89	2.21	4.45	2.17	4.29	1.50	4.05	0.92
(2)	3.75	3.13	3.63	2.53	3.50	2.14	3.40	1.51
(3)	3.11	5.14	3.00	5.08	3.27	3.72	3.36	2.11
(4)	1.95	13.66	2.05	11.62	3.27	5.04	3.46	1.02
(5)	1.95	13.74	1.18	11.67	3.22	5.05	3.32	1.51

注:表中各栏左边数据为拟合 RMSE, 右边数据为预测 RMSE。

表3 单项预测误差均值

预测	数 据 段							
	1~14	15~34	1~18	19~34	1~22	23~34	1~26	27~34
HCF	-1.00	2.10	-0.38	2.18	0.48	1.46	0.82	0.84
LBS	-0.75	1.18	-0.54	1.42	0.01	1.06	0.33	0.56
NI	-1.78	1.73	-1.39	2.17	-0.57	1.85	-0.10	1.56
OECD	-2.14	1.66	-1.08	1.41	-0.37	0.95	-0.04	0.54
PD	0.09	1.90	0.73	1.63	3.22	0.79	1.38	0.42

表4 单项预测方差

预测	数 据 段							
	1~14	15~34	1~18	19~34	1~22	23~34	1~26	27~34
HCF	5.51	5.75	3.25	5.78	4.55	5.26	6.48	4.88
LBS	4.36	5.19	4.50	5.58	5.39	5.36	6.06	5.00
NI	4.78	8.67	3.26	9.08	4.64	8.61	4.96	8.11
OECD	6.43	10.99	3.31	10.78	4.60	10.44	5.71	10.07
PD	5.25	4.19	4.06	4.02	5.51	3.70	7.63	3.63

3 结 束 语

本文通过实例分析,说明了当序列模式变动较大时,简单平均组合预测模型相对于其他组合预测模型的优越性,也说明了对历史数据依赖程度高的复杂组合预测模型难以适应序列模式大的变动。同时指出,当存在多变序列模式时,基于样本段的拟合精度或模型假设检验不足以说明模型的外推预测性能。本文的分析对于组合预测模型的选择和应用具有重要的实际参考价值。进一步的工作应考虑能适应多种序列模式的组合预测模型,文献[14]已取得了一定的进展。此外,简单平均组合的一个突出特点是其组合权重的高度稳定性,而这是适应性组合预测方法所不具备的。因此,能利用简单平均组合的权重鲁棒性和时变环境下的良好性能,又能对结构性变化作适应性调整的自适应组合预测方法值得进一步研究。

参 考 文 献

- 1 Dawes R, Filds R, Lawrence M *et al.* The past and the future of forecasting research. *Int J Forecasting*, 1994, 10: 151~159
- 2 Chatfield C. Neural network: forecasting breakthrough or passing fad? *Int J Forecasting*, 1993, 9: 1~3
- 3 Refenes A N. Comments on 'Neural networks: forecasting breakthrough or passing fad' by C Chatfield. *Int J Forecasting*, 1994, 10: 43~46
- 4 Chatfield C. Model uncertainty and forecast accuracy. *J Forecasting*, 1996, 15(7): 495~508
- 5 Makridakis S, Winkler R. Averages of forecasting: some empirical results. *Management Sci*, 1983, 20(9): 987~996
- 6 Bunn D W. Forecasting with more than one model. *J Forecasting*, 1989, 8(3): 161~166
- 7 唐小我. 预测理论及其应用. 成都: 电子科技大学出版社, 1992
- 8 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. *控制与决策*, 1994, 9(1): 20~28
- 9 曾 勇, 唐小我, 曹长修. 非负权重最优组合预测方法研究. *管理工程学报*, 1995, 9(3): 153~161
- 10 高仁祥, 张世英, 刘 豹. 组合预测贝叶斯方法研究. *系统工程学报*, 1996, 11(1): 28~35
- 11 Holden K, Peel D A. Unbiasedness, efficiency and the combination of economic forecasts. *J Forecasting* 1989, 8(3): 175~188
- 12 曾 勇, 唐小我. 几种无偏组合预测模型的分析. *数量经济技术经济研究*, 1996, (11): 41~45
- 13 Granger C W J, Ramanathan R. Improved methods of combining information. *J Forecasting*, 1984, 3(2): 197~204
- 14 Deutsch M, Granger C W J, Terasvirta T. The combination of forecasts using changing weights. *Int J Forecasting*, 1994, 10: 47~57

An Empirical Study of Efficiency of Averages of Forecasting

Zeng Yong

Li Yudong

(Management College, UEST of China Chengdu 610054) (Inst. of Command Automation of Navy Beijing 100036)

Tang Xiaowo

(Management College, UEST of China Chengdu 610054)

Abstract The advantages of averages of forecasting relative to other combining forecasts are illustrated by an empirical study. It is pointed out that the high accuracy of fitting is not sufficient for the high accuracy of combination forecasting when the mode of economic series is highly variable. The analysis in this paper is of importance to the selection and application of combination forecasting models.

Key words combination forecasting; averages of forecasting; accuracy of fitting; forecast accuracy; model uncertainty