

基于组合预测包容性检验的单项模型选择*

王丰效

(喀什师范学院 数学系, 新疆 喀什 844000)

摘要:组合预测模型实际上是单项预测模型的信息进行选择利用的过程。最优组合中单项预测模型的选择是十分重要的问题,将包容性检验应用于组合预测单项预测模型的选择;给出了基于组合模型包容性检验的单项模型的选择方法和步骤,该方法主要是利用增加一个单项预测方法的组合预测模型与原组合预测模型之间的包容性检验,确定单项模型是否要保留在组合预测模型中,这样可以达到提高组合预测的预测有效度。最后通过应用实例的分析,表明通过组合预测模型的包容性检验筛选出合适的单项预测模型,再建立组合预测模型就能够达到提高预测精度的效果,因此该方法是可行的和有效的。

关键词:组合预测;单项模型选择;包容性检验

中图分类号:O212.1

文献标志码:A

文章编号:1672-6693(2013)03-0081-04

在经济现象的预测中,通常对同一预测问题可以采用不同的预测方法建立不同的单项预测模型,不同的单项预测模型提供了不同的信息,预测精度一般也不同。组合预测将各种预测效果进行总体性综合,比单项预测模型更系统、更全面。组合预测的难点是最优模型组的选择和权重的分配,最优模型组包含的单项预测模型并非越多越好,很多文献研究了组合预测模型权系数的确定或优化方法^[1-6]。文献[7]提出了一个选择单项模型预测还是组合模型预测的判断方法,并提出了非线性组合预测方法,但同时指出如何确定最优模型组所包含的单项模型是尚未解决的问题。组合预测模型实际上是将单项预测模型的信息进行选择利用的过程,而研究组合预测中单项预测模型选择的比较少。在以往的组合预测研究及应用中,参与组合预测的单项预测模型往往没有经过筛选,而是主观地根据数据的特征和已有的经验选择单项预测模型,直接建立组合预测模型。本文将包容性检验的方法应用于组合预测模型,提出了两个组合预测模型的包容性检验方法,并将该检验方法应用于组合预测单项模型的选择中。首先利用灰色关联度对单项预测模型进行排序,然后对全部单项模型的组合预测模型和不包含预测性能最差单项模型的组合预测模型进行包容性检验,进一步判断是否剔除预测性能最差单项模型。最后给出了应用实例,说明了本文方法的有效性和可行性。

1 组合预测模型的包容性检验

1.1 组合预测模型

假设 $\{x_t, t=1, 2, \dots, n\}$ 为某经济现象的实际观测数据序列, x_{it} 为第 i 种单项预测方法 f_i 在第 t 个时刻的预测值,其预测误差 $e_{it} = x_t - x_{it}$,第 i 种预测方法 f_i 的预测误差向量为 $(e_{i1}, e_{i2}, \dots, e_{in})^T, i=1, 2, \dots, m, t=1, 2, \dots, n$ 。这 m 种单项预测方法的预测误差平方和以及两两之间的误差的协方差分别为 $E_i^T E_i = E_{ii}$ 和 $E_i^T E_j = E_{ij}, i, j=1, 2, \dots, m$ 。假定按照这 m 种单项预测方法建立组合预测模型,为了讨论方便,建立线性组合预测模型

$$x_t = l_1 x_{1t} + l_2 x_{2t} + \dots + l_m x_{mt}, t=1, 2, \dots, n \quad (1)$$

其中 $L = (l_1, l_2, \dots, l_m)^T$ 为权系数向量且满足 $\sum_{i=1}^m l_i = 1, l_i \geq 0, i=1, 2, \dots, m$ 。记 \hat{x}_t 为组合预测模型在第 t 个时刻的预测值,其预测误差 $e_t = x_t - \hat{x}_t$,误差预测平方和为 $J = L^T A_m L$,其中 $A_m = (E_{ij})_{m \times m}$ 称为误差信息矩阵。关于组合预测模型(1)权系数向量的确定方法有很多,可以利用均方误差准则,平均绝对相对误差等准则,也可以利用某种相关性指标,如预测有效度、灰色关联度、相关系数等来确定组合权系数。

非负权重组合预测问题中,有些单项预测方法的加入并不能降低非负权重组合预测模型的预测效果,这类方法就是冗余方法。在实际应用中应该把冗余方

* 收稿日期:2012-03-20 网络出版时间:2013-05-20 18:04

资助项目:国家社会科学基金(No. 11XTJ001)

作者简介:王丰效,男,教授,研究方向为随机序及其应用、统计组合预测, E-mail:fxw-hz@126.com

网络出版地址: http://www.cnki.net/kcms/detail/50.1165.N.20130520.1804.201303.81_015.html

法从组合预测结构中剔除,也就是要对参加组合预测的各种单项预测方法进行筛选。如果能将所有的冗余方法都剔除,就可以把组合预测的问题大大简化。

1.2 组合预测包容性检验

预测包容性检验的思想可以追溯到文献[8],用于确定某一预测是否包含其它预测中的有关信息。后来文献[9]进行了扩展,给出了预测包容性检验的回归模型。将预测包容性检验应用于组合预测中单项预测模型的选择,目前相关研究^[8-10]主要局限于两种单项模型组合时的预测包容性检验。在有多个模型进行组合的预测包容性检验中,也是通过两两包容性检验来进行模型选择,然后才进行组合的,但这样忽视了多个模型组合后可能包含组合前所不包容的单个模型的情况,组合模型中可能仍然包含冗余模型。基于以上不足,下面给出组合预测模型之间的包容性检验方法。

设 f_{ct} 为所有 m 种单项预测方法所组成的组合预测模型在第 t 个时刻的预测值,即

$$f_{ct} = \alpha_1 x_{1t} + \alpha_2 x_{2t} + \dots + \alpha_m x_{mt}$$

其中 $\alpha_1 + \alpha_2 + \dots + \alpha_m = 1, \alpha_k \geq 0, k = 1, 2, \dots, m$, 而设 $f_{(c-i)t}$ 为仅不含第 i 种单项预测方法所组成的组合预测模型在第 t 个时刻的预测值。考虑 f_{ct} 和 $f_{(c-i)t}$ 构成的回归模型

$$x_t = \beta_1 f_{(c-i)t} + \beta_2 f_{ct} + u_t \quad (2)$$

其中 $\beta_1 \geq 0, \beta_2 \geq 0, \beta_1 + \beta_2 = 1$ 为回归系数, u_t 为随机扰动。如果 $\beta_1 = 0, \beta_2 = 1$, 则称组合预测模型 f_{ct} 包容组合预测模型 $f_{(c-i)t}$, 如果 $\beta_1 = 1, \beta_2 = 0$, 则称组合预测模型 $f_{(c-i)t}$ 包容组合预测模型 f_{ct} 。如果 β_1, β_2 取其它值就称组合预测模型 f_{ct} 和 $f_{(c-i)t}$ 互不包容, 则表明单项预测方法 f_i 能够提供有用的信息。记 $e_{ct} = x_t - f_{ct}, e_{(c-i)t} = x_t - f_{(c-i)t}$, 分别表示两个组合预测误差, 则由(2)式可得

$$\begin{aligned} x_t - f_{(c-i)t} &= \beta_1 f_{(c-i)t} + \beta_2 f_{ct} + u_t - f_{(c-i)t} = \\ &= \beta_2 (f_{ct} - f_{(c-i)t}) + u_t = \beta_2 (e_{(c-i)t} - e_{ct}) + u_t \\ e_{(c-i)t} &= \beta_2 (e_{(c-i)t} - e_{ct}) + u_t \end{aligned} \quad (3)$$

这样组合预测模型的包容性检验问题就归结为对回归模型(3)的检验。回归模型(3)的检验步骤如下。

1) 提出假设 $H_0: \beta_2 = 0, H_1: \beta_2 \neq 0$, 在 H_0 为真时有

$$t = \frac{\hat{\beta}_2 - \beta_2}{S(\hat{\beta}_2)} = \frac{\hat{\beta}_2}{S(\hat{\beta}_2)} \sim t(n-2)$$

2) 计算统计量 t 的值并和临界值比较, 如果 $|t| \geq t_{\alpha/2}(n-2)$, 则拒绝原假设 H_0 , 表明单项预测方法 f_i 显著地增加了组合预测的精度。如果 $|t| < t_{\alpha/2}(n-2)$, 则接受原假设 H_0 , 表明单项预测方法 f_i 的增加没有改变组合预测的性能, 即组合预测模型 f_{ct} 和 $f_{(c-i)t}$ 的预测效果一样好, 单项预测方法 f_i 没有增加新的信息, 从而就可以将单项预测方法 f_i 删除。

当参加组合的预测方法由 $n+1$ 种减少到 n 时并

不能保证一定会增加非负权重组合预测的效果。非负权重组合预测问题中, 有些预测方法的去掉并不降低非负权重组合预测的效果, 因此组合预测中存在着冗余模型。

1.3 单项预测模型逐步遴选方法

组合预测的根本目的是通过在已有模型中包含更多的模型以改进预测的效果。也就是说, 当带有这个模型时的组合预测好于不带有这个模型时的组合预测时, 这个模型就应该包含在组合预测模型中, 否则, 就不应该将这个模型包含在组合预测模型中。下面在组合预测包容性检验方法的基础上, 给出基于组合预测模型包容性检验的单项预测模型选择方法。为了提高组合预测模型中单项预测方法选择的效率, 先对各个单项预测模型进行冗余性判定, 冗余预测方法直接剔除。另外, 为了减少检验次数, 可以先对各个单项预测方法的预测效果进行排序。组合预测单项模型选择的过程如下。

步骤 1 按照预测效果(如均方误差, 预测有效度, 灰色关联度等)对每个单项预测模型的预测性能进行评价, 并根据评价结果对模型进行排序。不妨假定各个单项预测方法的预测精度效果排序为 $f_1 > f_2 > \dots > f_m$ ($f_1 > f_2$ 表示模型 f_1 优于 f_2)。

步骤 2 先利用这些单项预测模型 f_1, f_2, \dots, f_m 建立组合预测模型 f_c , 再利用不包含预测精度最差的所有单项预测方法 f_1, f_2, \dots, f_{m-1} 建立组合预测模型 f_{c-m} 。对两个组合预测模型进行包容性检验, 如果组合模型 f_{c-m} 包容组合模型 f_c , 表明单项预测模型 f_m 的加入并没有使预测精度提高, 从而将单项预测模型 f_m 从单项模型组中剔除。否则, 则转入步骤 3。用 $f_{c-(m-1)}$ 代替 f_{c-m} 继续进行包容性检验。

步骤 3 用 f_{c-m} 代替步骤 2 中的 f_c , 与不包含单项预测模型 f_{m-1}, f_m 组合预测模型继续进行包容性检验, 直到所有的单项预测模型都进行了检验为止。

步骤 4 将最后得到的组合预测模型作为最终预测模型。

2 实例分析

例 1 以文献[10]中某单位舰船装备维修费的前 7 组数据为样本, 分别建立指数平滑法 f_1 , 灰色预测模型 f_2 , 参数法预测模型 f_3 , 指数法预测模型 f_4 , 二次多项式法 f_5 , 三次多项式法 f_6 , RBF 网络模型 f_7 , 这 7 种单项模型的预测值以及原始数据实际值见表 1^[10]。

为了简化计算, 先对 7 种单项预测模型按照预测误差平方和排序, 它们的预测误差平方和见表 1, 这样按照预测误差平方和, 这些单项预测模型的预测精度排序为

$$f_1 < f_3 < f_6 < f_7 < f_4 < f_5 < f_2$$

为了叙述方便, f_{123} 表示由 f_1, f_2, f_3 建立的组合

表 1 某单位舰船装备维修费实际值及预测值

序号	实际值	指数平滑	灰色模型	参数法	指数法	二次多项式	三次多项式	RBF
1	104.9	108.2	104.9	90.0	107.8	106.6	104.3	102.7
2	128.1	109.1	129.1	122.1	126.7	127	130.2	125.8
3	154.6	143.3	150.4	154.3	148.8	150.4	152.2	158.0
4	174.3	182.5	175.3	182.1	174.9	176.7	174.9	173.3
5	202.2	206.4	204.3	211.7	205.4	206	202.8	199.1
6	240.7	236.7	238.1	252.9	241.4	238.1	240.4	240.4
7	276.1	281.3	277.6	272.9	283.6	273.3	292.3	265.2
SSE		100.4	5.48	94.09	17.14	9.24	39.44	18.79

预测模型, f_{2567} 表示由 f_2, f_5, f_6, f_7 建立的组合预测模型, 组合预测模型的组合权系数都按照预测误差平方和倒数法确定。先用所有单项预测模型建立组合预测模型 $f_{1234567}$, 再建立仅不包含精度最差单项模型的组合预测模型, 即组合模型 f_{234567} 。对两个组合预测模型 $f_{1234567}$ 和 f_{234567} 按照 1.2 中的方法进行包容性检验(显著性水平为 0.05, 下同), 计算结果为 $p=0.1668 > 0.05$, 故接受原假设, 即可以认为组合模型 f_{234567} 包容了组合模型 $f_{1234567}$, 即单项预测方法 f_1 没有提供有效信息, 将 f_1 从单项模型组中删除。

继续检验组合模型 f_{24567} 和 f_{234567} , 计算结果 $p=0.6616 > 0.05$, 表明组合模型 f_{24567} 和 f_{234567} 预测效果相同, 因而删除单项模型 f_3 。继续检验组合模型 f_{24567} 和 f_{2457} , 计算结果 $p=0.0006 < 0.05$, 表明组合模型 f_{2457} 不包容组合模型 f_{24567} , 因而单项模型 f_6 提供了有效信息, 应该予以保留。检验组合模型 f_{24567} 和 f_{2456} , 计算结果 $p=0.0243 < 0.05$, 表明组合模型 f_{2456} 不包容组合模型 f_{24567} , 因而单项模型 f_7 提供了有效信息, 应该予以保留。

继续检验 f_{2467} 和 f_{24567} , 结果表明应该删除 f_5 , 检验 f_{267} 和 f_{2467} , 结果表明应该删除 f_4 。最后再继续检验, 计算结果表明, f_{26}, f_{27} 和 f_{67} 都不包容组合模型 f_{267} , 因此组合模型 f_{267} 就是最优组合预测模型。

上面通过逐步检验淘汰了被包容的一些单项预测模型, 本例中最后只要选择 f_2, f_6 和 f_7 这 3 种单项预测方法建立组合预测模型就能够达到比较好的预测效果, 按照这 3 种单项模型建立的组合预测模型为

$$\hat{x}_t = 0.699x_{2t} + 0.097x_{6t} + 0.204x_{7t}$$

为了进一步说明模型选择结果的合理性, 下面利用预测效果评价指标对几种组合预测结果进行分析比较。表 2 给出其中组合模型 $f_{1234567}, f_{24567}, f_{267}, f_{27}$ (分别记为组合模型 1, 组合模型 2, 组合模型 3 和组合模型 4) 的拟合结果, 这几种组合模型的绝对误差结果见图 1。

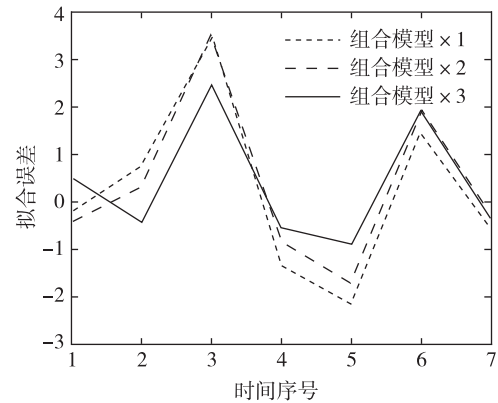


图 1 几种组合预测模型绝对误差比较

表 2 几种组合预测模型预测值及预测精度

序号	实际值	组合模型 1	组合模型 2	组合模型 3	组合模型 4
1	104.9	105.105 7	105.311 4	104.393	104.402 8
2	128.1	127.343 9	127.768 9	128.533 5	128.354 2
3	154.6	151.136 4	151.080 6	152.125	152.117 6
4	174.3	175.646 2	175.152 9	174.853 2	174.848
5	202.2	204.372 2	203.942 1	203.093 7	203.124 8
6	240.7	239.258 2	238.731 6	238.792 3	238.619 8
7	276.1	276.674 6	276.380 2	276.496 3	274.797 6
平均绝对误差		1.457 8	1.300 8	1.023 8	1.155 6
误差平方和		21.605	20.381	11.472	13.653
平均绝对相对误差		0.008 3	0.007 4	0.005 9	0.006 3

从表 2 的计算结果可以看出, 全部单项预测方法构成的组合预测模型的精度还没有其中 5 个单项预测方法构成的组合预测模型 f_{24567} 高, 而组合模型 f_{24567} 的精度也没有组合模型 f_{267} 高。但是仅由 2 个单项预测方法组成的组合模型 f_{27}, f_{26} 和 f_{67} 的预测精度都没有组合模型 f_{267} 高。从图 1 也可以看出组合模型 f_{267}

的绝对误差要比其它组合模型低。

上例也说明了组合预测模型并不是参与的单项预测模型越多越好。虽然这些单项预测模型的精度都不低, 但通过组合预测模型的包容性检验筛选出合适的单项预测模型, 再建立组合预测模型就能够达到提高预测精度的效果。

3 结论

本文在组合预测方法研究成果的基础上,对组合预测的单项模型的遴选问题进行了研究和讨论,利用组合预测模型之间的包容性检验给出了组合预测模型中单项预测模型选择的相应步骤和方法。在以往的组合预测研究及应用^[11-12]中,参与组合预测的单项预测模型往往没有经过筛选,而是人为主观决定后直接进行组合。而本文提出了判定参与组合预测的单项预测的方法,有效提高了组合预测精度。

参考文献:

- [1] 唐小我,曾勇. 组合预测误差校正应用模型的应用分析[J]. 管理科学学报,2002,5(6):53-64.
Tang X W, Zeng Y. Application of error correction models of combination forecasting[J]. Journal of Management Sciences in China, 2002, 5(6): 53-64.
- [2] 王丰效. 组合 GM(1,1) 幂模型及其应用[J]. 数学的实践与认识, 2011, 41(20): 124-129.
Wang F X. Combination GM(1,1) power model and its application[J]. Mathematics in Practice and Theory, 2011, 41(20): 124-129.
- [3] 唐小我. 组合预测误差信息矩阵研究[J]. 电子科技大学学报, 1992, 21(4): 448-454.
TANG X W. Study of combination forecasting error information matrix[J]. Journal of University of Electronic Science and Technology of China, 1992, 21(4): 448-454.
- [4] 陈华友, 赵佳宝, 刘春林. 基于灰色关联度的组合预测模型的性质[J]. 东南大学学报, 2004, 34(1): 130-134.
Chen H Y, Zhao J B, Liu C L. Properties of combination forecasting model based on degree of grey incidence[J]. Journal of Southeast University, 2004, 34(1): 130-134.
- [5] 陈华友. 基于相关系数的优性组合预测模型研究[J]. 系统工程学报, 2006, 21(4): 353-360.
Chen H Y. Research on properties of superior combined forecasting models based on correlation coefficients[J]. Journal of Systems Engineering, 2006, 21(4): 353-360.
- [6] 汤少梁, 李南, 巩在武. 灰色绝对关联度组合预测模型的性质研究[J]. 系统工程与电子技术, 2008, 30(1): 89-92.
Tang S L, Li N, Gong Z W. Research on properties of combination forecasting model based on absolute of grey incidence[J]. Systems Engineering and Electronics, 2008, 30(1): 89-92.
- [7] Yu L A, Wang S Y, Lai K K, et al. Time series forecasting with multiple candidate models: Selecting or combining[J]. Journal of Systems Science and Complexity, 2005, 18(1): 1-18.
- [8] Chong Y Y, Hendry D F. Econometric evaluation of linear macroeconomic models[J]. Review of Economic Studies, 1986, 53(4): 671-690.
- [9] Fang Y. Forecasting combination and encompassing tests[J]. International Journal of Forecasting, 2003, 19(1): 87-94.
- [10] 谢力, 魏汝祥. 基于包容性检验的舰艇装备维修费组合预测[J]. 系统工程与电子技术, 2010, 32(12): 2599-2602.
Xie L, Wei R X. Combined forecasting of ship equipment maintenance cost based on encompassing tests[J]. Systems Engineering and Electronics, 2010, 32(12): 2599-2602.
- [11] 曾波. 基于改进灰色预测模型的电力需求预测研究[J]. 重庆师范大学学报: 自然科学版, 2012, 29(6): 99-104.
Zeng B. Research on electricity demand forecasting based on improved grey prediction model[J]. Journal of Chongqing Normal University: Natural Science, 2012, 29(6): 99-104.
- [12] 王丰效, 周伟萍. 灰色多元线性回归方法的改进及应用[J]. 重庆理工大学学报: 自然科学版, 2012, 26(8): 113-116.
Wang F X, Zhou W P. Improvement of grey multiple linear regression analysis and application[J]. Journal of Chongqing University of Technology: Natural Science, 2012, 26(8): 113-116.

Stepwise Selection Method of Single Model for Combination Forecast Based on Encompassing Test of Combination Models

WANG Feng-xiao

(Dept. of Mathematics, Kashi Teachers College, Kashi Xinjiang 844000, China)

Abstract: Combination forecasting model is actually the use of the individual prediction models. How to select the single models in combination forecast is very important. The encompassing test method of combination forecasting models is applied to select single models. The single model group selection method is presented based on encompassing test of combination forecasting models to improve precision of combination forecasting model. The method is mainly to determine whether a single model to be retained in the combined forecasting model according to encompassing test between combined forecasting models of a single forecast method increased with the original combination forecasting model. Finally, the analysis of the application example shows that the predictive accuracy effect of combination forecasting model is improved according to the combined forecasting model encompassing test select single models, so that this method is feasible and effective.

Key words: combination forecasting model; single model select; encompassing test

(责任编辑 游中胜)