

ARIMA 模型、BP 神经网络及其组合模型在卫生政策评估中的实证比较:以公立医院价格改革为例

马爱霞* 谢 静 唐文熙

中国药科大学国际医药商学院 江苏南京 211198

【摘要】目的:探索不同反向事实构建方法对医院财务数据预测的效率,以期对政策进行更有效的评估。方法:借助 R 软件,用南京市公立医院 A 在 2011—2016 年的药品收入、医疗服务收入建立测试数据集,分别用 ARIMA 模型、BP 神经网络、ARIMA + BP 组合模型进行预测并与实际拟合,并比较改革前后补偿率。结果:三个模型对药品收入的均方根误差分别为 692.82、501.44、380.80,医疗服务收入的均方根误差分别为 184.04、215.63、168.65,组合模型预测效率更高。用组合模型计算改革后 A 医院药品收入净损失为 12 044.03 万元,医疗服务收入净增长为 18 532.60 万元,为药品收入损失的 153.87%。结论:医院财务数据因其线性与非线性的组合特征,使用组合预测模型的预测效果最佳。但在实际应用中,ARIMA 模型操作简单,与组合模型预测趋势也较为一致,在实际卫生政策评估中也推荐使用。

【关键词】ARIMA 模型; BP 神经网络; 组合模型; 药品收入; 医疗服务收入; 政策评价
中图分类号:R197 文献标识码:A doi:10.3969/j.issn.1674-2982.2018.01.012

A comparison of ARIMA Model, BP Neural Network Model and combined model in health policy evaluation: An empirical study of public hospitals pricing reform

MA Ai-xia, XIE Jing, TANG Wen-xi

School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing Jiangsu 211198, China

【Abstract】 Objective: To study the effectiveness of different time series models in the prediction of financial data in public hospitals, with the aim of obtaining a more reliable counterfactual in health policy evaluation. Methods: ARIMA model, BP neural network and their combination were used for the estimation and prediction of drug revenue and medical service revenue based on a dataset for the period from November, 2011 to October, 2016 for hospital X before and after Nanjing medical pricing reform. Root mean square error (RMSE) was used to estimate the model accuracy. Results: RMSE of drug revenue from the three models were 692.82, 501.44 and 380.80, and of medical service were 184.04, 215.63 and 168.65. The findings shows that the combination model was proved to be the most efficient one among the three. The combined model was used to calculate the net loss of drug revenue which was estimated to be 120,440 million, and the net increase of medical service was estimated to be 185,326 million after the reform, which was 1.539 times of the drug loss. Conclusions: The revenue data of public hospitals are usually complex with a both linear and non-linear trend. The combination model of ARIMA and BP neural network could solve the problem for once with an acceptable accuracy. However, ARIMA model is simpler to operate as compared to other two models, and also more consistent with the forecasting trend, therefore ARIMA is also recommended in the evaluation for health policies.

【Key words】 ARIMA model; BP neural network; Combined model; Drug revenue; Medical service revenue; Policy evaluation

* 基金项目:国家自然科学基金青年项目(71603278);江苏省教育厅哲学社会科学基金(2016SJD630007)

作者简介:马爱霞,女(1961年—),教授,博士生导师,主要研究方向为药物经济学与医药产业政策。

E-mail: ma86128@sina.com

通讯作者:唐文熙。E-mail: tokammy@cpu.edu.cn

城市公立医院医药价格综合改革后,由于取消药品加成、提高医疗服务价格,医院药品收入与医疗服务收入将会受到影响,其中医疗服务收入的改变更是直接反映了改革取得的效果。一般来说,随着社会经济的发展,医疗服务收入本身呈增长趋势,因此在评价政策效果时,应该剔除数据本身存在的规律,即构造“反事实”——通过自身历史数据构建其发展趋势并用以预测“非改革”状态下数据的表现情况。医院收支数据多为时间序列,常用的预测模型有线性时间序列模型、ARIMA 模型 (Autoregressive Integrated Moving Average Model)、BP (Back Propagation) 神经网络模型等。^[1-3] 本文选用 ARIMA 模型、BP 神经网络模型以及 ARIMA + BP 组合模型,探讨其预测效率,以期为政策评价中反事实的构造提供方法学参考。

1 模型简介

1.1 ARIMA 模型

ARIMA 模型全称为差分自回归移动平均模型 (Autoregressive Integrated Moving Average Model), 是时间序列模型最常用的参数模型之一, 由 Box 和 Jenkins 于 20 世纪 70 年代初提出。^[4] 模型运算过程为: 先将非平稳时间序列转化为平稳时间序列, 然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归。ARIMA 模型由于在逼近许多平稳过程时具有较高的适应性, 因此常用于线性预测。模型完整形式为 ARIMA(p, d, q), 其中 AR 是自回归, p 为自回归项; MA 为移动平均, q 为移动平均项数, d 为时间序列成为平稳时所做的差分次数。平稳化后预测公式如下所示:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} + \mu_t - \Theta_1 \mu_{t-1} - \Theta_2 \mu_{t-2} - \dots - \Theta_q \mu_{t-q} \quad (1)$$

针对许多时间序列数据波动具有季节规律性, 可将 ARIMA 模型扩展为季节时间序列模型 (Seasonal ARIMA Model, SARIMA), 即本研究采用的模型。SARIMA 模型采用季节差分的方法去掉季节性, 可解决 ARIMA 模型无法解决的季节性问题。季节性建模处理与 ARIMA 模型类似。模型完整形式为 SARIMA(p, d, q)(P, D, Q)^s, 公式表达式如下所示:

$$\varphi_p(L)\varphi_p(L^s)(1-L)^d(1-L^s)^D X_t = \Theta_q(L)\Theta_q(L^s)\varepsilon_t \quad (2)$$

其中, P 为季节自回归阶数, Q 为季节移动平均阶数, s 为周期, $\varphi_p(L^s)$ 和 $\Theta_q(L^s)$ 分别为季节 P 阶自

回归算子和 Q 阶移动平均算子。E_t 服从均值为 0 方差为 σ^2 的正态分布。建立 SARIMA 模型, 应首先对季节性因素进行处理, 用观察法或统计检验确定季节性后, 对时间序列进行季节平稳性检验, 然后识别 SARIMA 模型, 确定 P, D, Q, p, d, q 的取值, 随后进行参数估计与模型检验, 得到初步模型后即可使用 SARIMA 模型进行时间序列的仿真与分析。

1.2 BP 神经网络模型

BP (Back Propagation) 神经网络, 即误差反向传播模型, 由信息的正向传播和误差的反向传播两个过程组成, 是一种最为常见且较为复杂的神经网络模型。^[5] BP 神经网络是一个三层或以上的网络, 不仅包含输入层和输出层, 而且还有一层或多层隐层 (图 1)。X = (X₁, X₂, ..., X_n) 即为 BP 神经网络的输入向量, b = (b₁, b₂, ..., b_t) 是隐含层的输出向量, Y 是输出层的输出向量, W_{ih} 和 W_{hj} 是 BP 神经网络的网络权值。

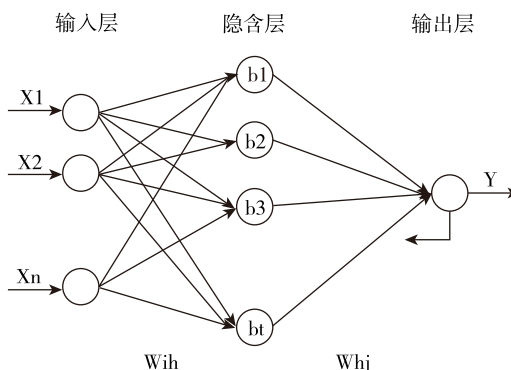


图 1 三层 BP 神经网络结构图

此时节点的输出模型为:

$$b_h = f_1 \left(\sum_{i=1}^{N_1} w_{ih} x_i \right), i = 1, 2, \dots, N_1 \quad (3)$$

$$y_j = f_2 \left(\sum_{h=1}^{N_2} w_{hj} y_h \right), h = 1, 2, \dots, N_2 \quad (4)$$

式(3)是隐含层节点的模型, 其中 f₁ 为激活函数, 一般取输出值在 (0, 1) 之间的 Sigmoid 函数。式(4)是输出层节点的模型, 其中 f₂ 为激活函数, 通常为线性函数。当实际输出与期望输出不符时, 进入误差的反向传播阶段。误差通过输出层, 按误差梯度下降的方式修正各层权值, 向隐藏层、输入层逐层反传。周而复始的信息正向传播和误差反向传播过程, 是各层权值不断调整的过程, 也是神经网络学习训练的过程, 此过程一直进行到网络输出的误差减少到可以接受的程度, 或者预先设定的学习次数

为止。

BP神经网络模型对单变量的时间序列也可进行预测。以某一时间序列 X_1, X_2, \dots, X_n 为例,预测时则认为预测值与前面 m 个值之间存在某种函数关系,构建 $X_t = F(X_{t-1}, X_{t-2}, \dots, X_{t-m+1})$ 函数关系。即将预测值 X_t 当为输出变量 Y ,该值前面的 m 个值($X_{t-1}, X_{t-2}, \dots, X_{t-m+1}$)作为输入向量 X ,利用BP神经网络来推导未来值。近年来随着人工智能的不断发展,BP神经网络已经成为预测非线性时间序列的主流模型之一。

为解决时间序列的季节性波动问题,从一般的BP神经网络发展出了季节性神经网络模型。设一时间序列 $X = (X_1, X_2, \dots, X_n)$ 是以 s 为周期的季节性时间序列,根据季节性ARIMA模型理论,认为未来的一个序列 X_t 值是和历史 $d \times s$ 个值之间存在某种函数关系,函数关系为 $X_t = F(X_{t-1}, X_{t-2}, \dots, X_{t-d \times s+1})$,用三层神经网络(图1)拟合该周期函数,来预测未来的 X_t 值。

1.3 ARIMA + BP 组合模型

组合预测模型的思想最早在1969年由Bates等提出,Bates等多位研究学者均指出并验证了单项预测模型的有效组合可以显著地提高预测效果。^[6]将ARIMA与BP预测模型进行组合,能够结合ARIMA与BP模型的优势,分别提取研究数据的线性和非线性变化特征。^[7]目前使用最广泛的是线性组合预测模型,一般的线性组合预测模型在时间序列应用时表达式如下:

$$y_t = w_1 y_{t1} + w_2 y_{t2}$$

其中 y_t 为组合模型的预测结果, y_{t1}, y_{t2} 分别为ARIMA和BP模型预测模型 t 期的预测值, w_1, w_2 为相应的组合权重。组合模型的难点在于最优组合权重的确定。若权重不合理,则有可能出现组合模型的预测效果比单项模型差的情形。目前,确定权重的方法中被大多学者认可并采用的为数学优化方法,即通过求误差平方和的最小值来确定最优的权重系数。假设ARIMA、BP模型、组合模型的预测误差分别为 e_{1t}, e_{2t}, e_t ,欲使组合模型误差平方和 $E = \sum e_i^2$ 最小,则有:

$$e_t = w_1 e_{1t} + w_2 e_{2t}$$

$$w_1 + w_2 = 1$$

将 w_2 用 $1 - w_1$ 代入可以得到:

$$e_t = w_1 (e_{1t} - e_{2t}) + e_{2t}$$

组合模型误差平方和为:

$$E = \sum e_i^2 = \sum [w_1 (e_{1t} - e_{2t}) + e_{2t}]^2$$

通过计算 E 的最小值,即可得此时 w_1, w_2 的值(即最优权重)。

2 资料与方法

2.1 资料来源

本研究从南京市城市公立医院医药价格综合改革中选取代表性医院A,并选取药品收入、医疗服务收入两个指标进行预测。A医院为三级甲等医院,改革前后药占比分别为40.91%和36.09%,医疗服务收入占比分别为16.63%和20.92%,在57所参改医院中属于中等,且规模适当、管理规范,因此较有代表性。

2.2 研究方法

鉴于南京价改于2015年10月31日零点正式开始,样本区间设定为2011年11月—2016年10月。按月收集改革前4年及改革后1年的数据,以改革前3年(2011年11月—2014年10月)的数据为训练数据,以改革前1年(2014年11月—2015年10月)为测试数据,分别运用ARIMA模型、BP神经网络及其组合模型对指标进行预测,并与未构造反事实情况下的指标表现水平进行比较。数据分析过程在R软件中实现。

3 结果

3.1 药品收入预测比较

建模前对原始数据进行预处理,拟分析的数据形态如图2所示。可见,数据有一定的上升趋势,同时又伴随有周期性约为12个月的季节性波动,符合医院药品月收入现实意义。

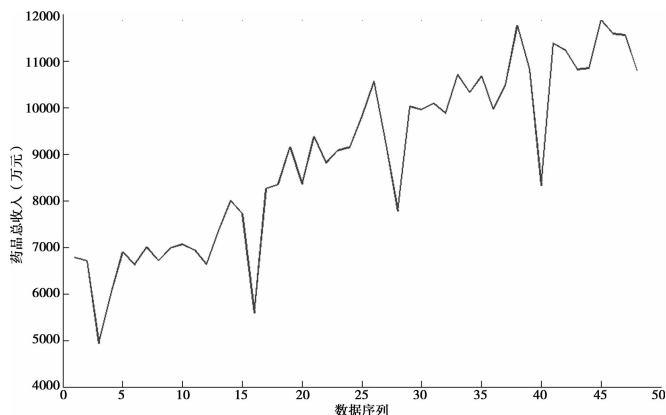


图2 药品收入原始数据

SARIMA 模型:在 R 软件中先调用 ts() 函数将原始数据转化为波动周期为 12 的季节性时间序列,后运用 auto.arima() 函数,判断通过一次差分和季节差分后数据转化为一个平稳时间序列,再根据药品月收入时间序列一阶差分的自相关和偏自相关函数构建模型为 ARIMA(0,1,1) × (1,1,0)₁₂。

BP 神经网络模型:本研究采用滞后项预测法,所构建神经网络模型的输入层节点数为 12,输入的是待预测数据前的 12 个时间序列数据 $y_{t-1}, y_{t-2}, \dots, y_{t-12}$;输出层节点数为 1,输出的是待预测数据 y_t 。此外,隐含层的节点数通过经验法与试凑法确定为 10,最终神经网络模型结构为 12-10-1。隐含层的激活函数为 log-sigmoid 函数,输出层激活函数为线性函数,设定迭代次数为 1 000 次,学习率为 0.02,误差目标为 0.00001。考虑到本文的数据预

测属于时间序列长期预测,因此采用迭代一步预测的方法。

组合模型:借助软件计算得到 ARIMA 模型权重 $w_1 = 0.3895$,BP 模型权重 $w_2 = 0.6105$ 时,此时组合预测模型的误差平方和最小。

三种模型预测结果数值如表 1 所示,选用均方根误差(RMSE)对预测效果进行评估。指标表达式如下:

$$RMSE = \sqrt{n^{-1} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

从表 1 可看出组合模型的均方根误差最小,预测效果最为精准;三者的预测结果趋势如图 3 所示,其中测试数据是指 2014 年 11 月—2015 年 10 月测试数据的真实值,可以发现组合模型的拟合趋势与测试数据最为接近,也进一步说明了其预测效果的精准性。

表 1 三种模型对药品收入预测结果比较(万元)

月份	实际药品收入	药品收入预测值			模型预测误差		
		ARIMA 模型	BP 神经网络	组合模型	ARIMA 模型	BP 神经网络	组合模型
11	10 482.39	10 935.28	10 163.58	10 499.83	-452.89	318.81	-17.44
12	11 758.98	11 625.77	11 652.89	11 840.84	133.21	106.09	-81.85
1	10 819.45	10 893.25	10 043.36	10 102.74	-73.79	776.09	716.72
2	8 326.66	9 059.75	9 073.24	9 146.75	-733.09	-746.58	-820.09
3	11 381.52	11 555.54	11 703.62	11 619.41	-174.02	-322.10	-237.89
4	11 233.69	11 580.60	10 627.89	10 746.42	-346.91	605.80	487.27
5	10 818.43	12 101.02	10 489.27	10 554.99	-1 282.60	329.16	263.44
6	10 840.79	11 547.19	10 340.00	10 492.48	-706.40	500.79	348.31
7	11 879.06	12 494.13	11 364.12	11 439.92	-615.07	514.94	439.14
8	11 586.58	12 000.43	11 330.00	11 357.21	-413.85	256.59	229.38
9	11 552.26	12 303.77	11 708.27	11 743.08	-751.51	-156.01	-190.82
10	10 796.27	12 055.33	10 048.25	10 167.97	-1 259.06	748.01	628.30
均方根误差	—	—	—	—	692.82	501.44	380.80

3.2 医疗服务收入预测比较

同理,对医疗服务收入进行测算。医疗服务收入 = 医疗收入—药品收入—卫材收入—检查收入—化验收入。组合模型的最优系数分别为: $w_1 = 0.6658$; $w_2 = 0.3342$ 。计算结果如表 2,发现组合模型预测的准确度最高。三个模型预测结果见图 4。

3.3 与实际值比较

3.3.1 改革对药品收入的影响

根据前文研究的结果,组合模型预测效果最优,但为了对比不同模型预测与组合模型预测的差值情况,本文分别采用组合模型、ARIMA 模型以及 BP 模型对改革后 A 医院的药品收入进行反向事实数据预算。如图 5 所示,三种模型预测趋势较为接近,并且发现改革后预测药品收入与实际药品收入相比有明显下降。

表2 三种模型对医疗服务收入预测结果比较(万元)

月份	实际医疗服务收入	医疗服务收入预测值			模型预测误差		
		ARIMA	BP神经网络	组合预测	ARIMA模型	BP神经网络	组合模型
11	4 181.36	4 464.09	4 092.19	4 339.80	-282.73	89.17	-158.44
12	4 835.80	4 765.45	4 570.18	4 700.19	70.35	265.62	135.61
1	4 208.32	3 817.69	4 108.11	3 914.75	390.63	100.21	293.57
2	4 009.31	3 977.11	4 148.98	4 034.55	32.20	-139.68	-25.24
3	4 794.27	4 502.33	4 517.55	4 507.42	291.94	276.72	286.85
4	4 491.62	4 412.96	4 266.90	4 364.15	78.65	224.72	127.47
5	4 610.70	4 583.59	4 711.71	4 626.41	27.10	-101.01	-15.71
6	4 550.31	4 603.35	4 752.75	4 653.28	-53.03	-202.43	-102.96
7	5 009.29	4 792.30	4 618.06	4 734.07	217.00	391.24	275.23
8	4 732.19	4 744.99	4 965.78	4 818.78	-12.80	-233.59	-86.59
9	4 553.83	4 620.29	4 783.01	4 674.67	-66.46	-229.18	-120.84
10	4 497.17	4 643.17	4 390.06	4 558.58	-146.00	107.11	-61.41
均方根误差	—	—	—	—	184.04	215.63	168.65

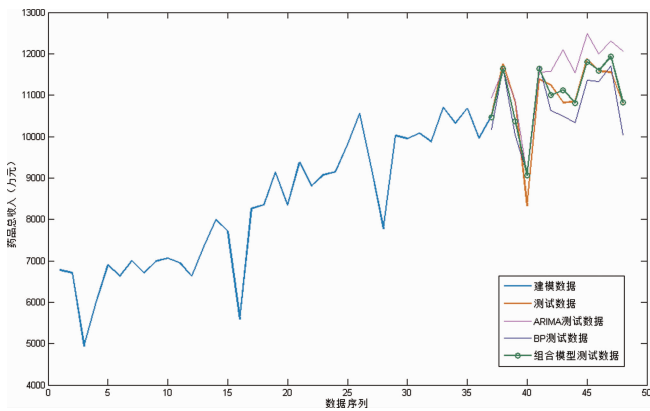


图3 药品收入预测结果

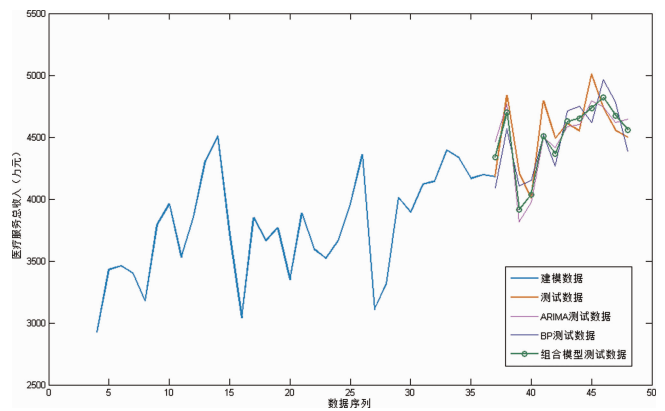


图4 医疗服务收入预测结果趋势图

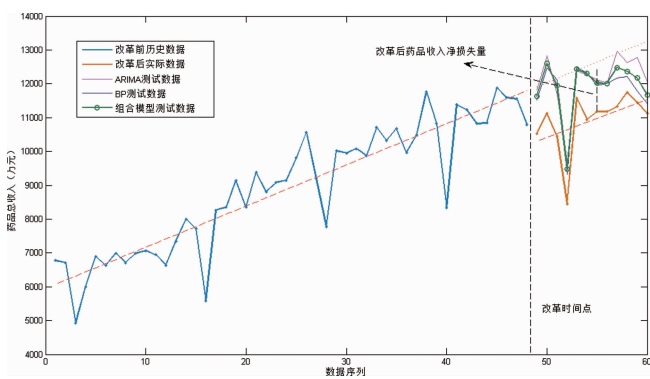


图5 改革后药品收入损失情况

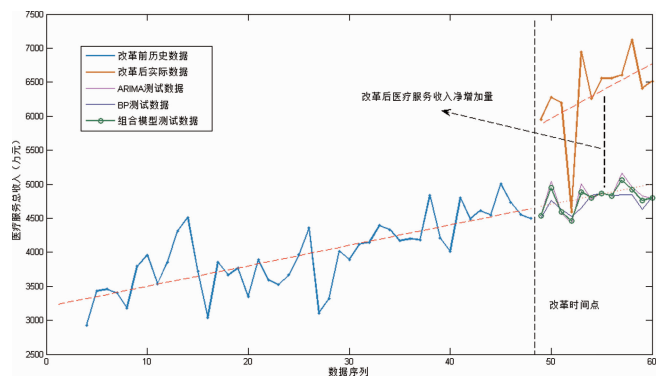


图6 改革后医疗服务收入增加情况

三种模型预测的反向事实数据如表3所示,用改革后1年实际药品收入减去改革前1年实际药品收入,获得传统的改革药品收入损失量为403.95

万元,由于药品收入自身具有上升的趋势,未剔除自身发展趋势,将导致药品收入损失预测降低。运用反向事实数据预测得到的改革后预测药品收入,

利用改革后实际药品收入减去改革后预测药品收入,得到 ARIMA、BP 神经网络、组合模型三种模型预测的药品收入净损失量分别为 14 180.53 万元、10 680.94 万元、12 044.03 万元。相比于传统计算获得的药品收入损失,三种运用反向事实数据预测获得的净药品收入损失更贴近改革实际带来的净效应。

3.3.2 改革对医疗服务收入的影响

同理,对医疗服务收入分别进行组合模型、ARIMA 模型以及 BP 模型进行反向事实数据预算。如图 6 所示,三种模型预测的趋势均较为接近,并且发现改革后实际医疗服务收入比预测医疗服务收入有明显上升。

表 3 改革后药品收入反向事实数据的预测(万元)

月份	改革后实际药品收入	改革前药品收入	改革后预测药品收入		
			ARIMA 模型	BP 神经网络	组合模型
11	10 524.54	10 482.39	11 815.61	11 516.19	11 632.82
12	11 129.58	11 758.98	12 832.53	12 469.32	12 610.79
1	10 453.27	10 819.45	11 680.27	12 125.14	11 951.87
2	8 448.52	8 326.66	9 717.76	9 332.28	9 482.42
3	11 572.28	11 381.52	12 372.06	12 468.08	12 430.68
4	10 946.95	11 233.69	12 264.83	12 319.84	12 298.42
5	11 175.42	10 818.43	12 119.87	11 952.38	12 017.61
6	11 176.02	10 840.79	12 027.76	12 002.42	12 012.29
7	11 328.05	11 879.06	12 963.61	12 176.48	12 483.07
8	11 749.04	11 586.58	12 624.75	12 202.67	12 367.07
9	11 447.55	11 552.26	12 782.05	11 773.69	12 166.45
10	11 120.91	10 796.27	12 051.57	11 414.58	11 662.69
总和	131 072.13	131 476.08	145 252.67	141 753.08	143 116.17
药品收入损失值	—	403.95	14 180.53	10 680.94	12 044.03

三种模型预测的反向事实数据如表 4 所示,按照传统计算方式:改革后医疗服务收入—改革前医疗服务收入,获得医疗服务收入增加量为 22 541.3 万元,由于医疗服务收入自身具有上升的趋势,未剔除自身发展趋势,将导致医疗服务收入增加量预测加大。运用反向事实数据预测得到的改革后医疗服

务收入净增加量,ARIMA、BP 神经网络、组合模型三种模型预算的结果分别为 18 172.25 万元、19 250.51 万元、18 532.60 万元。相比于传统计算获得的医疗服务收入增加,三种运用反向事实数据预测获得的净医疗服务收入增加量更贴近改革实际带来的净效应。

表 4 改革后医疗服务收入反向事实数据的预测(万元)

月份	改革后实际医疗服务收入	改革前医疗服务收入	改革后预测医疗服务收入		
			ARIMA 模型	BP 神经网络	组合模型
11	5 952.80	4 181.35	4 551.67	4 507.59	4 536.94
12	6 276.31	4 835.80	5 037.49	4 762.99	4 945.75
1	6 197.63	4 208.32	4 571.69	4 629.53	4 591.02
2	4 592.00	2 976.66	4 423.95	4 524.93	4 457.70
3	6 941.73	4 794.27	5 006.66	4 648.15	4 886.85
4	6 258.23	4 491.61	4 781.99	4 832.57	4 798.89
5	6 555.75	4 610.69	4 870.38	4 862.76	4 867.84
6	6 558.68	4 550.31	4 825.56	4 828.86	4 826.66
7	6 606.48	5 009.29	5 166.28	4 841.54	5 057.75
8	7 120.70	4 732.18	4 960.57	4 844.56	4 921.80
9	6 408.86	4 553.82	4 828.17	4 629.20	4 761.67
10	6 513.60	4 497.16	4 786.11	4 819.57	4 797.29
总和	75 982.82	53 441.52	57 810.57	56 732.31	57 450.21
医疗服务收入增加值		22 541.30	18 172.25	19 250.51	18 532.60

3.3.3 医疗服务补偿率

医疗服务补偿率是指取消药品加成后,医院通过调整医疗服务价格获得的对药品差价补偿的比例。传统的医疗服务补偿率和剔除数据自身发展趋势以获得改革后的医疗服务净补偿率计算公式如下:

传统医疗服务补偿比 = 传统的改革医疗服务收入增加量/传统的改革药品收入损失量 * 100%

医疗服务收入净补偿比 = 改革后医疗服务收入净增加量/改革后药品收入净损失量 * 100%

根据表 3、表 4 分别计算出传统医疗服务补偿

比、与运用反向事实 3 个模型测算的医疗服务补偿比(表 5)。可以发现,按照传统计算方式得到的结果与运用反向事实预测得到的结果相差较大,这是由于药品收入和医疗服务收入自身均具有上升的趋势,未剔除自身发展趋势,将导致药品收入损失预测降低、医疗服务收入增加量预测加大,因此计算出的医疗服务补偿率与实际情况偏差较大。观察剔除数据自身发展趋势后获得的医疗服务净补偿率,可以发现补偿率已经超过 100%,证明改革后该医院增加的服务收入能够全额弥补甚至超出药品价格差额损失。

表 5 医疗服务补偿率

指标	改革后实际值	改革前实际值	反向事实预测		
			ARIMA 模型	BP 神经网络	组合模型
药品收入(万元)	131 072. 13	131 476. 08	145 252. 67	141 753. 08	143 116. 17
服务收入(万元)	75 982. 82	53 441. 52	57 810. 57	56 732. 31	57 450. 21
医疗服务收入补偿比(%)		5 580. 22	128. 15	180. 23	153. 87

4 讨论

4.1 反向事实的重要性

从前文的分析中不难看出反向事实构建对医疗政策评价的准确度具有重要意义。现阶段,我们对医疗卫生政策的研究多采用对比改革前后观测指标的方法来衡量改革成效^[8-9],但这种评价方法存在一定的局限性。其中最主要的问题是忽视了指标数据自身发展趋势(即反向事实数据)对效果评价的影响,在指标数据自身呈现正向或负向趋势时,这种忽视就直接导致了改革后观测指标的变化因素除了改革的影响外还包括了其自身客观发展的影响。具体而言,当指标数据具有正向发展趋势时,忽视反向事实数据将会导致评价机会成本下降、机会收益增加,同样,忽视具有反向发展趋势的反向事实数据则会导致评价机会成本上升、机会收益下降。因此,剔除数据自身发展趋势即将反向事实数据纳入评价体系对获得改革的净效应具有重要意义。运用反向事实评价公立医院价格综合改革成效将最大限度还原改革所获得的净效应。

4.2 组合模型捕捉线性和非线性效率更高

ARIMA 模型与大多数时间序列预测方法一样,各变量之间的线性关系也是其主要预测前提之一^[10],但实际研究中总是或多或少地含有非线性因素,当非线性因素影响较小,或在某一范围内影响较

小时,尚可采用线性模型来描述或逼近。但非线性影响较大或用线性逼近也得不到较好结果的时候,非线性时间序列模型的运用就显露了其特有的优势。近年来神经网络模型的兴起为非线性模型预测提供了新思路。由于具备自动学习、逼近能够反映样本数据规律的最优函数,且当函数形式越复杂时,神经网络预测的效果越好^[11],其在预测高复杂度的非线性时间序列方面明显优于传统的线性预测方法。

总结而言,ARIMA 模型可对具有线性关系的数据进行时间序列预测,对非线性数据的处理不尽合理,且效果欠佳;BP 神经网络可挖掘出数据中隐含的非线性关系,但由于不能反映时间序列的自相关和偏相关,在处理具有线性特征的数据时,其效果往往不如 ARIMA。现实研究中,很多数据都是既包含线性信息,也包含非线性信息,单个的预测模型很难充分捕获时间序列中所包含的信息,而此时组合模型可以结合不同预测模型的优势,尽可能多的获取时间序列数据中的信息。

医院财务数据是医院医疗工作的经济基础、健康发展的基石,更是评价公立医院综合改革成效的重要依据。由于医院财务数据会受到各类因素波动的影响,如人口结构、疾病谱变化、季节性流行疾病、医生处方习惯、药品价格、服务价格等,均会导致医院财务数据具有复杂的线性与非线性组合特

征,因此在用组合预测模型的预测效果最佳。

4.3 时间序列模型的局限性及选择建议

在实际操作时, BP 神经网络模型隐含层层数和单元数的选择在理论上尚缺乏有力指导,一般是根据经验或反复实验进行确定,一定程度上增加了网络学习的负担。网络的学习和记忆还具有不稳定性,其对之前的权值和阈值缺乏记忆,在增加学习样本的情况下,网络需从头训练,故而 BP 及其组合模型的预测也会比较难。^[12] ARIMA 模型预测精度虽然没有组合模型预测的效果好,但预测的趋势与组合模型趋势也较为一致,且其操作性简单,实际应用性相对较强。

一般而言,在预测时间序列数据时,当数据长度越长,模型拟合预测效果也将越好,但本研究的时序数据仅限于十三五规划期间的数据,建议其他学者在有条件的基础上选取更长的数据进一步研究探索。

作者声明本文无实际或潜在的利益冲突。

参 考 文 献

- [1] 李望晨,崔庆霞,张利平. 基于趋势外推与 ARIMA 预测我国医院诊疗及住院人次[J]. 中国卫生统计, 2016, 33(3): 477-478.
- [2] 侯福均,吴祈宗. BP 神经网络在铁路客运市场时间序列预测中的应用[J]. 运筹与管理, 2003, 12(4): 73-75.
- [3] 翟静,曹俊. 基于时间序列 ARIMA 与 BP 神经网络的组

合预测模型[J]. 统计与决策, 2016(4): 29-32.

- [4] 于爽,郭祖超,胡琳. ARIMA 模型在医院季节性时序资料预测中的应用[J]. 数理统计与管理, 1991(6): 23-29.
- [5] 杨娟丽,徐梅,王福林,等. 基于 BP 神经网络的时间序列预测问题研究[J]. 数学的实践与认识, 2013, 43(4): 158-164.
- [6] 梁德阳. 基于 SARIMA 和 BP 神经网络的时间序列组合预测模型研究[D]. 兰州:兰州大学, 2014.
- [7] 张宇青,易中懿,周应恒. 一种线性 ARIMA 基础上的非线性 BP 神经网络修正组合方法在粮食产量预测中的运用[J]. 数学的实践与认识, 2013, 43(22): 135-142.
- [8] 肖月,赵琨,李雪,等. 东部某省县级公立医院医药价格改革评价[J]. 中国卫生政策研究, 2015, 8(1): 14-20.
- [9] 陕西省价格协会陕西省物价局课题组. 城市公立医院补偿机制改革研究[J]. 价格理论与实践, 2016(10): 72-77.
- [10] 叶明全. 基于神经网络的季节性时间序列预测方法研究[D]. 合肥:合肥工业大学, 2004.
- [11] 钟颖,汪秉文. 基于遗传算法的 BP 神经网络时间序列预测模型[J]. 系统工程与电子技术, 2002, 24(4): 9-11.
- [12] 杨伟,倪黔东,吴军基. BP 神经网络权值初始值与收敛性问题研究[J]. 电力系统及其自动化学报, 2002, 14(1): 20-22.

[收稿日期:2017-07-17 修回日期:2017-10-18]

(编辑 刘博)