

# 基于 BP 神经网络的股价趋势预测

——以房地产开发上市公司为例

朱永明(教授) 邵庚云

(郑州大学管理工程学院 郑州 450001)

**【摘要】** 本文从上市公司的财务影响因素角度出发,以房地产开发业为例,采用 BP 神经网络预测公司股价。首先,通过采用主成分分析法优化输入变量,并且得出主成分的回归方程;然后,将训练好的网络保存并且采用实证数据进行预测;最后,将回归方程法和 BP 神经网络法预测精度进行比较,结果表明,BP 神经网络法的预测精度明显高于回归方程法。

**【关键词】** 股价趋势 主成分分析法 回归方程法 BP 神经网络法

## 一、问题的提出

股票市场作为我国金融市场中的重要组成部分,对国家经济水平的发展影响越来越大。股票投资作为一种收益较高的直接融资手段也受到各类机构和人们的青睐。然而,我国股票市场自 2008 年遭受重创后,仅在 2009 年年初小幅反弹之后就一蹶不振,我国证券市场总市值也从 2007 年的 6.23 万亿美元大幅缩水到 2008 年的 2.79 万亿美元,上证指数也从 2007 年的 6 000 点坠入 2012 年年底的 1 949 点,股价变动可谓跌宕起伏。

Fama(1970)提出有效市场理论,认为资本市场的个性业务活动就像毫无规律可循的布朗运动,市场对投资者是公平的,任何人都不能窃取超额收益。但是,大量的实证数据表明,资本市场不是完全有效的,我国的股票市场属于弱有效的资本市场。国外多位学者认为,价格没有反映所有公开信息,股价的波动前后也是有关系的,股票价格的分布经常表现出偏峰厚尾的特征,收益率不再符合随机游动规律。从这一角度看,能够找寻一种描述股票市场变动规律特征的模型,建立相应的预测方法在遵循规律的基础上进行预测是关键所在。

学者们对国内外股价预测方法的研究也先后涌现,最初是多元线性回归模型等传统的模型方法,还有一些常见的方法如时间序列分析法,包括传统的移动平均与分解模型、指数平滑模型。近几年出现的比较现代的模型有 ARIMA 模型、随机波动模型(SV)等。人工神经网络最早可追溯到 1943 年 McCulloch 和 Walter Pitts 发表的论文,1987 年 Lapedes 等人最先把神经网络用于预测领域。20 世纪 90 年代初,Matsuba 提出人工神经网络可以用于预测股票价格。我国应用神经网络是从廖晓昕(1994)的研究开始取得新的进展。

近几年来,我国学者在神经网络进行股价预测方面也有一些研究。如张立军、苑迪(2008)基于遗传算法,提出一种 GA-Elman 动态回归神经网络的股价预测模型。李婷婷

(2008)为了提高金融股票价格预测的准确性,采用一种改进的 BP 神经网络建立时间序列预测模型。许兴军(2011)采用 BP 神经网络进行股价趋势分析并且深入分析股价涨幅情况。

综合来看,以前学者研究选取的初始变量大多集中在历史价格和成交量这两个因素上,同时影响股价的诸多因素中,过于倾向技术层面的指标分析,难以反映影响股价变化的其他重要因素,因而在预测准确性上仍然有上升空间。

我国的大多数散户投资者偏爱跟风炒作,盲目炒题材股、概念股,关注政策层面多于技术层面,更加忽视对上市公司微观财务状况的分析。他们在预测方法上采用简单的市盈率法分析股价的实际价值。市盈率是根据当前股价与每股收益的比值计算得到,再与行业市盈率进行对比,看是否低于行业市盈率,如果低则代表被低估,反之则被高估。但是,考虑到行业标准市盈率不具有代表性,因此采用市盈率法判断股票的价值是不科学的。

本文提出一种 BP 神经网络的预测方法对次日股价进行预测,以帮助投资者判断该股票是否被低估。同时,在数据变量预处理时采用主成分分析找出各个财务指标的主成分,避免变量之间的非独立性影响预测结果。同时,本文将 BP 神经网络和传统的回归分析进行精度对比,以突出神经网络的精确性。

## 二、BP 神经网络模型

人工神经网络(ANN)是运用到生物界中人脑的神经元传递信号,具有学习记忆功能,以此来模拟人脑处理大量信息,实现多输入多输出的数据并行处理优点。前向反馈(BP)网络和径向基(RBF)网络是神经网络中最为成熟的两种技术。通常,BP 神经网络是利用梯度学习法的 RBF 网络的特殊形式。本文拟通过利用 BP 神经网络对一定时间内影响某上市公司股价变动因素的变量进行预处理后的主成分作为输入变量,对次日股价进行预测。

BP神经网络是一种具有至少三层的神经网络，输入层、中间层、输出层上都有不同数量的神经元，两层之间的神经元通过网络权值进行连接。中间层和输出层上的神经元还会通过加上阈值的形式对前一层神经元传递过来的信息进行整合。神经元之间学习会朝误差减小的方向传递，从而修正了各自的权值，这种算法即为误差反向传播算法，也称为BP算法。误差减小的方法采用的是负梯度下降法，即误差沿着下降的方向总是朝下降最快的方向进行。

$$\text{隐含层的输出为: } y_{ij} = f\left(\sum_i w_{ij}x_i - \theta_j\right)$$

$$\text{输出层节点的输出为: } y_{jk} = f(y_j) \left[ \sum_j w_{jk} f\left(\sum_i w_{ij}x_i - \theta_j\right) - \theta_k \right]$$

$$\text{网络的作用函数为 Sigmoid 函数: } f(x) = \frac{1}{1 + e^{-x}}$$

由于激励函数在接近0和1时收敛速度减慢，因此将训练样本、测试样本以及网络预测的输入样本值置于0.1-0.9之间，具体做法为： $X^* = 0.1 + 0.8 \frac{X - X_{\min}}{X_{\max} - X_{\min}}$ 。

$$\text{若 } T_{jk} \text{ 为期望输出值, 则误差: } E = \frac{1}{2} \sum_i (T_{jk} - y_{jk})^2;$$

$$\text{输出层节点误差为: } e_j = T_{jk} - y_{jk};$$

$$\text{隐含层节点误差为: } e_k = f'(y_{jk}) \sum_i e_j w_{jk}。$$

BP神经网络最为重要的是网络拓扑结构的搭建和学习速率的取值大小。网络层一般为三层，较复杂的学习样本为四层，本文实证研究采用三层BP网络，并且采用“隐节点合成”的方法。BP算法的步骤如下：①初始化：确定输入输出层为三层，确定隐含层节点数、随即初始权值、学习系数、动量系数，合成时刻误差  $\varepsilon_1$ ，目标误差  $\varepsilon_2$ ，阈值  $\theta_1$ 、 $\theta_2$ ，令当前的学习次数为  $k=1$ 。②BP学习：用BP算法规则对神经网络的误差进行修正，同时计算当前训练的误差值和对所有模式的隐节点输出序列。③合成、终止判断：如果当前的训练误差值  $E > \varepsilon_1$ ， $k=k+1$ ，转到步骤2； $\varepsilon_2 < E < \varepsilon_1$ ，转到步骤4； $E < \varepsilon_2$ ，终止学习。④计算各隐节点输出序列的标准差和相关系数。⑤合成相关性较高的隐节点。如果隐节点  $i$  和  $j$  相关系数满足  $|r_{ij}| > \theta_1$ ，且  $s_i^2 > \theta_2$ 、 $s_j^2 > \theta_2$ ，则对隐节点进行合成。⑥偏移节点合成，如果某隐节点满足  $s_i^2 < \theta$ ，则对该隐节点同偏移点进行合并， $k=k+1$ ，转到步骤2。

### 三、输入量的预处理

1. 指标的建立。股票市场行情波动极大，首先受到国内外政治、经济、军事等各种宏观因素的影响，同时也受到市场外财政政策和货币政策（诸如经济周期、汇率、利率、通货膨胀率、货币供应量、政策倾向热点）等因素影响。其次，市场内部各种投资机构的交易资金变动、供求关系与市场需求、证监会、银监会等监管部门的政策指向性以及市场发展规律，这些因素同样会引发股票市场的震荡。再次，上市公司内部财务和

非财务因素（如经营业务的现金流营运能力、盈利能力、偿债能力、发展能力、股东变动、产权变动、利润分配等因素）同样会对股价产生影响。

表 1

市场外因素(宏观因素)		市场因素	公司因素
政治	国际间利益争端等	投资机构变动	主业营运能力
经济周期	财政政策、货币政策	市场供求关系	盈利、偿债、发展能力
军事	战争等	监管制度	股东变动
社会	行业动向	市场基础	利润分配

在众多影响股价的因素中，除去宏观市场内外因素等系统性风险之外，最能够得到控制和调整的是公司因素。如果企业能够提升公司的盈利水平、偿债水平、营运能力、发展能力，平衡好投资融资的资本结构，制衡好股东之间的利益分成等问题，公司因素完全可以正面积积极影响股价的波动。因此，公司财务和非财务的因素是影响股价上涨的重要部分，投资者在分析某只股票的投资价值时也要把分析公司因素放在首要位置。

在指标体系的构建中本文选择能反映房地产开发行业上市公司的反映其财务偿债能力、营运能力、盈利能力、发展能力的16大指标，具体说明见表2。利用SPSS19.0对指标进行主成分分析，从所选财务指标中提取主成分，并且对股价进行多元回归分析，将提取的主成分作为输入变量对股价进行预测。

表 2 财务指标体系构建

指标类型	指标代码	指标名称	指标计算说明
现金流与偿债能力	X1	流动比率	流动资产/流动负债
	X2	每股经营活动现金流量	经营活动现金净流量/总股数
	X3	资产负债率	总负债/总资产
	X4	现金流动负债比	经营活动现金净流量/流动负债
营运能力	X5	现金周转率	现金余额/平均资产总额
	X6	存货周转率	主营业务成本/平均存货余额
	X7	应收账款周转率	赊销收入净额/平均应收账款
	X8	流动资产周转率	主营业务收入净额/平均流动资产总额
盈利能力	X9	销售净利率	净利润/销售收入
	X10	营业利润率	营业利润/销售收入
	X11	净资产收益率	净利润/平均净资产
	X12	每股收益	净利润/总股数
	X13	每股净资产	所有者权益/总股数
成长能力	X14	总资产收入增长率	(本期总资产-上期总资产)/上期总资产
	X15	净资产增长率	(期末净资产-期初净资产)/期初净资产
	X16	净利润增长率	净利润/平均资产总额

2. 多元线性回归分析过程。主成分分析是通过降维将多个变量变成少数几个综合变量或主成分的数据预处理方法。本文通过对影响房地产业上市公司股价变动的16个财务指标进行降维，提取与之相关的几大主成分，并且对这几大主成

分拟合成回归方程进行解释。本文选取上证指数房地产开发业 50 家上市公司 2011 年年报的数据计算的指标体系进行主成分分析,得到的结构如下:

	成 分					
	1	2	3	4	5	6
VAR00008	0.968	-0.097	-0.071	0.045	-0.045	-0.005
VAR00014	0.954	-0.016	-0.009	-0.095	0.074	0.029
VAR00005	0.923	-0.058	-0.088	-0.099	0.022	-0.001
VAR00006	0.872	-0.100	-0.060	0.121	0.003	0.072
VAR00009	-0.122	0.953	0.125	-0.092	-0.030	0.014
VAR00010	-0.109	0.937	0.115	-0.217	0.042	0.024
VAR00012	-0.073	0.254	0.867	-0.037	0.075	-0.069
VAR00013	-0.100	0.232	0.788	0.054	0.020	-0.286
VAR00015	-0.044	-0.200	0.627	-0.049	-0.140	0.269
VAR00004	0.011	-0.146	0.010	0.918	0.024	-0.071
VAR00002	-0.010	-0.120	-0.017	0.879	0.109	0.067
VAR00011	-0.021	0.146	0.037	0.095	0.721	0.133
VAR00007	-0.003	-0.173	-0.137	0.091	0.656	-0.237
VAR00016	0.363	0.027	0.235	-0.406	0.544	0.291
VAR00003	-0.122	-0.125	-0.197	-0.074	0.254	0.798
VAR00001	-0.159	-0.134	-0.066	-0.047	0.167	-0.642
特征值	4.792	2.746	1.818	1.433	1.172	1.039
贡献率	29.948%	17.162%	11.361%	8.957%	7.327%	6.496%
累计贡献率	29.948%	47.11%	58.471%	67.429%	74.756%	81.252%

我们在将指标标准化之后,对标准化的指标进行主成分分析,得到各主成分的特征值和贡献率。我们取累积贡献率为 81.252%,特征值大于 1 的成分为主成分,即取前六个成分  $Z_1$ 、 $Z_2$ 、 $Z_3$ 、 $Z_4$ 、 $Z_5$ 、 $Z_6$  代替 16 个指标作为输入指标变量。由表 2 可以看出, $X_1$ 、 $X_9$ 、 $X_{10}$ 、 $X_{11}$ 、 $X_{12}$ 、 $X_{13}$ 、 $X_{15}$  与  $Z_1$  之间负相关,且载荷量较大, $X_5$ 、 $X_6$ 、 $X_8$ 、 $X_{14}$  皆为周转率的指标,因此  $Z_1$  主要是反映企业的运营能力; $Z_2$  中载荷量较大, $X_9$ 、 $X_{10}$  体现企业的盈利能力,因此  $Z_2$  反映企业的盈利能力。同理, $Z_3$  反映企业股东权益创造财富的能力; $Z_4$  反映企业现金流能力; $Z_5$  反映企业的投资回报能力; $Z_6$  反映企业偿债能力。我们选取这六大主成分作为 BP 神经网络的输入变量值。

选取主成分  $Z_1$ 、 $Z_2$ 、 $Z_3$ 、 $Z_4$ 、 $Z_5$ 、 $Z_6$  与 F 进行多元回归,结果如下:

$$F=0.876+0.054 4Z_1+0.059 63Z_2+0.612 8Z_3-0.070 57Z_4+0.048 21Z_5+0.0196 1Z_6$$

方程整体显著性水平平均低于 0.05,拟合度较高。

#### 四、BP 神经网络分析过程

1. 样本数据来源。本文的因变量 Y 以 2012 年 4 月 30 日的收盘价为基准,自变量以 2011 年年末财务报表为基准计算。选取上交所房地产开发行业 50 只股票的 2011 年年报进

行自变量计算,数据主要来自国泰安数据库、巨潮资讯网和同花顺股票软件。将 50 个样本数据中前 40 个作为网络训练样本,将后 10 个样本数据作为检验样本集。

2. 参数确定。如前文所述,输入变量通过主成分分析法的优化处理确定了六大因子  $Z_1$ 、 $Z_2$ 、 $Z_3$ 、 $Z_4$ 、 $Z_5$ 、 $Z_6$  作为输入变量,本文的 BP 神经网络初始参数采用 6-2-1 的三层拓扑结构,较小的学习系数为 0.1,较大的动量系数为 0.9。通过确定好的初始参数,我们可以输入数据对网络进行训练,按照顺序优选实验中需要的三个主要参数:隐含层节点数、动量系数  $\alpha$ 、学习系数  $\eta$ 。

隐含层节点数目太少会降低网络的分类能力,太多则造成无用学习。在  $\eta=0.1$ 、 $\alpha=0.9$  的初始参数下,我们将隐含层节点数的选取范围定为 1~6 进行实验。通过反复实验发现:当隐含层节点数为 2 时,训练和预测的平均误差均较大,说明隐含层节点数过少,网络不能很好地进行信息归类;当隐含层节点数逐步增加至 5 时,从实验中可以看到网络训练的时间延长,训练和预测的误差也明显增加;而当隐含层节点数为 4 时,网络训练的平均误差和预测平均误差都是最小的,迭代次数也在合理的范围,因而实验中将隐层节点数确定为 4 个。在隐含层节点数为 4、学习系数为 0.1 的前提下,取不同的  $\alpha$  进行实验。实验中发现,当  $\alpha$  取 0.5 或 0.6 时,迭代次数比较理想,训练误差和预测误差均比较小,因而追加  $\alpha$  取 0.55 的实验。从上表中可知,虽然迭代次数增加,但更重要的是网络的训练误差和预测误差均比较小,因而当  $\alpha$  为 0.55 时,网络的性能达到最佳。最后,令隐含层节点数为 4, $\alpha$  为 0.55,取不同的  $\eta$  进行实验。实验发现  $\eta$  为 0.25 时,训练误差和预测误差均比较小,网络性能达到最佳。

通过上述的参数确定,最终我们选择 6-4-1 的拓扑结构进行测算。同时,选择训练误差和预测误差均最小时的学习效率  $\eta=0.25$ ,动量因子  $\alpha=0.55$ 。本文设定的最大训练次数为 5 000 次,目标精度为 0.005,选取 2012 年 4 月 30 日作为交易日记录实验数据,随机选取 40 组数据进行 BP 网络训练,剩下的 10 组交易数据作为测试检验。

3. 神经网络的预测结果。将训练好的 BP 神经网络保存并将预处理好的  $Z_1$ 、 $Z_2$ 、 $Z_3$ 、 $Z_4$ 、 $Z_5$ 、 $Z_6$  作为输入变量进行预测,以 5 次实验预测结果的平均值作为最终预测结果。通过 Matlab 对随机选取的 40 组交易数据进行 682 次训练后,训练误差降到了 0.009 87,将 10 组测试检验数据输入进行预测检验,预测样本的迭代误差为 0.000 936,达到了误差要求。

通过预测检验,发现 BP 神经网络的预测精度远远高于线性回归模型。

表 4 两种方法精度对比

分析方法	训练样本精度(%)	检验样本精度(%)
多元回归模型	65.8	59.2
BP神经网络	98.65	81.2

# 中联重科应收账款质量分析

刘宏<sup>1</sup> 朱一鸣<sup>1</sup> 景舒婷<sup>2</sup>

(1.桂林电子科技大学商学院 广西桂林 541004 2.桂林电子科技大学信息科技学院 广西桂林 541004)

**【摘要】**企业极易利用应收账款的坏账计提比例操纵盈余管理。本文以中联重科2006~2011年的财务报表信息为依据,实证研究了其应收账款管理质量。研究表明:2011年中联重科在行业整体步入调整阶段时大幅下调了应收账款的坏账计提比例,虽然提高了账面利润,但掩盖了盈利质量。

**【关键词】** 工程机械 应收账款 坏账计提

2008年金融危机后,工程机械行业经过了一段高速增长时期,受产能增速过快和产业结构不合理的影响,2011年进入了调整和理性回归阶段。激烈的销售手段透支了部分市场需求,有奖销售、低首付、融资租赁比例过大、租赁公司门槛过低、小型农用工程机械享受政府补贴等竞争扰乱了行业秩序,非理性竞争导致企业的应收账款畸高、资金链紧张。

行业巨头之一中联重科2012年3月15日审议通过了《会计估计变更》,大幅下调了短期应收账款坏账计提比例。由于没有合理估计宏观经济政策变化带来的市场风险,因利率、

准备金不断上调使投资信贷环境趋紧,下游客户偿债能力下降带来的潜在损失,这些都不符合会计信息质量要求的谨慎性原则。2011年年报显示,该项会计估计变更增加了归属于上市公司股东的净利润16 132.40万元,占公司2011年度净利润的2%。我国对应收账款坏账计提比例没有强制性规定,法制上的不足为企业操纵盈余管理提供了空间。

## 一、行业视角下应收账款与主营业务收入的财务分析

2007年至2011年工程机械行业14家上市公司应收账款与主营业务收入数据见表1。

## 五、结论

诸多实证研究表明,我国股票市场的有效性较弱,所以在我国股票市场上进行股价的分析和预测在一定程度上是可行的。影响股价变动的因素很多,宏观政策层面、微观企业层面以及市场的技术层面都会影响股价变动,投资者的行为及投资心理也会影响股价变动。本文同以往经常出现的采用BP神经网络进行股价预测的研究不同,以往选取的输入指标大多是技术层面的指标,不够全面,不能综合反映影响股票市场变动的因素。同时,本文还分别采取人工神经网络和回归分析对我国房地产开发业50家上市公司股价进行预测,通过误差分析比较,发现BP神经网络预测的准确性更高。

但是,本文选取的输入指标仅仅是对反映微观层面的上市公司财务情况的指标进行训练和预测,对于影响因素众多、复杂多变的股票价格来说显然不足。在今后的研究过程中,将选取综合反映估计变动的指标,如宏观变量、外盘股市行情等指标作为输入变量来进行更加准确的预测。同时,股票市场系统是非常复杂的,其中广泛存在着非线性、时变性和不确定性。神经网络不一定是最为适用的预测模型,因此,未来还将在模型适用性方面进行研究,并且和多种预测方法结合起来,以寻找最优模型进行预测。

**【注】**本文系河南省社科调研课题“ERP环境下集团化企业财务管控体系研究”(项目编号:12-3375)的阶段性研究成果。

## 主要文献综述

1. 黄人杰. 中国资本市场监管有效性研究. 投资研究, 2010;7
2. 张根明,任福匀. 我国证券市场行业板块的波动特点分析. 统计与决策, 2006;1
3. 刘红梅. ARIMA模型在股票价格预测中的应用. 广西轻工业, 2008;6
4. 张立军,苑迪. 基于GA-Elman动态回归神经网络的股价预测模型研究. 华东经济管理, 2008;9
5. 李婷婷. 基于大数集模糊BP神经网络的金融股票市场的预测. 科技信息, 2008;11
6. 许兴军,颜钢锋. 基于BP神经网络的股价趋势分析. 证券保险, 2011;11
7. 张代远. 神经网络新理论与方法. 北京:清华大学出版社, 2006
8. 孙泉,赵旭峰,钱存华. 基于多点加权马尔可夫链模型的股价预测分析. 南京工业大学学报, 2008;5