DOI:10.19789/j.1004-9398.2017.03.003

基于移动平均线改进的多元非线性回归模型*

李馨旖 周 清
(北京邮电大学理学院,北京 100876)

摘 要

在技术分析中常用移动平均线预测股价未来走势,但普通移动平均线仅考虑收盘价和时间的关系,存在一定局限性,因此提出了对移动平均线的分步改进: 第一步,用成交额对移动平均线进行加权,得到成交额加权的移动平均线 VTMA; 第二步,借鉴指数移动平均值的计算公式对第一步的公式进行改进,得到成交额加权的类指数平滑移动平均线 VTEMA; 第三步,对第二步用三角移动平均值的计算方法进行双重平滑得到 DVTEMA. 为了预测股票次日收盘价,建立基于 DVTEMA、MACD、KDJ、RSI 的多元非线性回归模型.

关键词: 移动平均线, 成交额, 指数移动平均, 多元非线性回归模型.
中图分类号: O29

0 引 言

运用移动平均线预测股价的未来趋势在技术分析中占有一席之地,然而简单移动平均线仅考虑价格和时间的关系使其本身具有局限性[1]. 目前已有的改进是用成交量对移动平均线加权[2]. 假设近期的股票价格变化不大,早期股票价格波动较大,但是在计算中所占的权重较小甚至被去掉,导致"数据忽略效应". 为解决这一问题,采用新的计算方法,指数移动平均(Exponential Moving Average)[3]; 三角移动平均的本质是双重平滑,综合使用指数移动平均和三角移动平均并用成交额代替成交量对移动平均线进行加权. 对改进后的移动平均线用统计学的方法进行分析和解释. 最后,分别用多元线性回归模型[4]和多元非线性回归模型进行预测,对预测的结果进行分析和评估.

1 成交额加权的移动平均线

成交量是指股票一天的成交总量,即交易量和交易价格之积. 以 10 日移动平均线为例(MA10),通常的计算方法只与收盘价和周期有关. 通常的改进方法是用开盘价、最高价、最低价、收盘价的算术平均值代替简单移动平均线中的收盘价. 用 SPSS 对金融(浦发银行)、冶金(武钢股份)、环保(首创股份)、供电产业(华能国际)、车辆工程(东风汽车)、石油(中国石化)、机械(三一重工)、交通运输(白云机场)、传媒娱乐(歌华有线)、房地产(保利地产) 10 个主要行业的具有代表性的股票进行皮尔逊相关分析,计算数据为 2013 年 1 月 4 日至 2014 日 12 月 31 日,482 个交易日.

所有数据都能通过相关性检验. 并且, 10 只股票都有一个相同点:

$$\mathrm{corr}(\text{成交额},\mathrm{MA10}) > \mathrm{corr}(\text{成交量},\mathrm{MA10}). \quad (1)$$

其中,corr(成交额,MA10)为 10 日移动平均线值(MA10)与收盘价的皮尔逊相关系数. 改进方法之一是用$\frac{O_i + C_i + H_i + L_i}{4}$代替 C_i,其中 O_i – 第 i 日的开盘价[2]; C_i – 第 i 日的收盘价; H_i – 第 i 日的最高价; L_i – 第 i 日的最低价. 但不同股票之间存在差异,加之使用的是长期的 482 个连续的交易日的计算,若对于短期行情,简单的平均数可能就不足以体现灵敏性. 改进的方向是,对开盘价、最高价、最低价、收盘价进行动态加权平均,用它们与 MA10 的相关系数作为权重,每更新一天的数据就重新计算一次权重,计算的结果将用于下一个交易日 MA10 的计算. 公式如下:

收稿日期: 2016-06-11
* 国家自然科学基金(编号: 11471051, 11371362)资助项目.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\mathrm{Price}_i = [\mathrm{corr}(O_i, \mathrm{MA}\,10_{i-1})\,O_i + \mathrm{corr}(C_i, \mathrm{MA}\,10_{i-1})\,C_i + \mathrm{corr}(H_i, \mathrm{MA}\,10_{i-1})\,H_i + \mathrm{corr}(L_i, \mathrm{MA}\,10_{i-1})\,L_i] \div [\mathrm{corr}(O_i, \mathrm{MA}\,10_{i-1}) + \mathrm{corr}(C_i, \mathrm{MA}\,10_{i-1}) + \mathrm{corr}(H_i, \mathrm{MA}\,10_{i-1}) + \mathrm{corr}(L_i, \mathrm{MA}\,10_{i-1})]. \quad (2)$$

对比用 $\frac{O_i + C_i + H_i + L_i}{4}$ 近似替代收盘价的方法,上述方法可以更加体现每只股票的价格波动特点,从而计算 MA 时会更精确. 但为了后续的计算方便,在后续的方法中还是使用开盘价、最高价、最低价、收盘价的简单算术平均值作为计算移动平均线时的参考价格. 成交额加权的移动平均线计算公式如下:

$$\mathrm{VTMA} = \frac{\sum_{i=1}^{N} \mathrm{TV}_i \times \frac{(O_i + C_i + H_i + L_i)}{4}}{\sum_{i=1}^{N} \mathrm{TV}_i}, \quad (3)$$

其中 VTMA 为成交额加权的移动平均值,TV 为成交额. 皮尔逊相关分析的结果是成交额加权的移动平均值(VTMA)与次日股价的相关性略优于成交量加权的移动平均值(VMA),相关性越紧密意味着预测率的提高.

2 利用指数平滑移动平均线(EMA)的改进

2.1 问题描述

算术移动平均值和线性加权移动平均值会产生"数据忽略效应". 为了解决这个问题,采用指数移动平均值(EMA)的计算方法进行改进[4]. 其公式为:

$$\mathrm{EMA}_{\mathrm{today}} = \alpha \times \mathrm{price}_{\mathrm{today}} + (1 - \alpha) \times \mathrm{EMA}_{\mathrm{yesterday}}. \quad (4)$$

其中,平滑指数 $\alpha = \frac{2}{T+1}$,$\mathrm{price}_{\mathrm{today}}$ 为当日股票收盘价.

$$\mathrm{VTEMA}_{\mathrm{today}} = \alpha \times \mathrm{price}_{\mathrm{today}} + (1 - \alpha) \times \mathrm{VTMA}_{\mathrm{yesterday}}. \quad (5)$$

其中,VTEMA 为成交额加权的类指数平滑移动平均线. VTEMA 的计算方法与指数移动平均值(EMA)的算法形式上相似,但本质不同,EMA 的实质是递归算法,当递归到初始值时需要人为给定初始值;而 VTEMA 并没有用到递归,而是提高了当日股票收盘价价格的权重. 一般来说收盘价的参考意义最大.

2.2 VTEMA 的改进效果分析

取周期 $T = 10$,则平滑指数 $\alpha = \frac{2}{10+1} \approx 0.181\,8$,$(1 - \alpha) \approx 0.818\,2$;用 2013 年 1 月 4 日至 2013 年 11 月 7 日的浦发银行的股票数据. 从与次日股价相关性的角度总结,VTMA 比起 VMA 平均提高了 0.000 275 679,而 VTEMA 又比 VTMA 平均提高了 0.003 354 496,用 VTEMA 与 MA 直接对比:

$$\frac{\mathrm{corrVTEMA-10,次日股价} - \mathrm{corr(MA-10,次日股价)}}{\mathrm{corr(MA-10,次日股价)}} \times 100\% \approx 5.799\%, \quad (6)$$

可以看出 VTEMA 的改进效果非常明显,相关性越紧密意味着预测率的提高,将会更有利买卖时机的选择.

3 双重平滑的 VTEMA(DVTEMA)

3.1 DVTEMA 的计算

用双重平滑的思想对上一节中的 VTEMA 进一步改进得到双重平滑的基于成交额加权的类指数移动平均线(DVTEMA). 三角移动平均的核心思想是用同样的方法进行双重平滑,VTEMA 的计算参见公式(5).

DVTEMA 进一步加大了单日收盘价的权重,更适用于短期预测. 下面列表说明 DVTEMA 的计算过程(以浦发银行 2013 年 1 月 18 日至 2013 年 1 月 22 日为例):

表 1 DVTEMA 计算过程

A	B	C	F
交易日期	收盘价	VTEMA-10	DVTEMA-10
2013/1/18	10.16	10.124	
2013/1/21	10.4	10.172	10.174
2013/1/22	10.48	10.193	10.228

周期 $T = 10$,则平滑指数 $\alpha = \frac{2}{10+1} \approx 0.181\,8$,$(1 - \alpha) \approx 0.818\,2$;

F 列为 D 列加 E 列的和,则 2013 年 1 月 21 日的 DVTEMA-10 计算式如下:

$$\mathrm{DVTEMA} - 10 = \partial \times 10.4 + 1 - \partial \times 10.124 = 10.174. \quad (7)$$

3.2 DVTEMA 改进效果分析

从与次日股价相关系数的角度总结,DVTMA 比 VTEMA 平均提高 0.001 742 16,而 VTEMA 又比 MA 平均提高 0.051 082 176,DVTEMA 比起简单移动平

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

均线 MA 的相对改进效果为:

$$\frac{\text{corrDVTEMA}-10\text{,次日股价}-\text{corr}(\text{MA}-10\text{,次日股价})}{\text{corr}(\text{MA}-10\text{,次日股价})}\times 100\% = \frac{0.051\,082\,176}{0.850\,797\,35}\approx 6.004\%. \quad (8)$$

对比,式(8)和式(5),DVTEMA 比起上一节的 VTEMA 在预测次日股价的相关性方面又有所提高.

4　回归方程

4.1　一元线性回归模型

用 SPSS 做次日收盘价关于移动平均值的回归分析[5],用 2013 年 1 月 4 日至 2013 年 11 月 7 日的浦发银行的股票数据为例进行计算. 五个模型的回归模型如下:

MA 模型: $Y_1 = 0.042 + 1.166X_{\text{MA}}$, (9)

VMA 模型: $Y_2 = 0.792 + 0.916X_{\text{VMA}}$, (10)

VTMA 模型: $Y_3 = 0.809 + 0.913X_{\text{VTMA}}$, (11)

VTEMA 模型: $Y_4 = 0.675 + 0.927X_{\text{VTEMA}}$, (12)

DVTEMA 模型: $Y_5 = 0.587 + 0.937X_{\text{DVTEMA}}$, (13)

其中 X 是当日浦发银行的移动平均值, Y 是次日的收盘价.

用上述 5 个模型进行预测,因为模型的参数是用 2013 年 1 月 4 日至 2013 年 11 月 7 日的浦发银行的股票数据进行拟合的,那么将各模型中移动平均值在 2013 年 11 月 7 日的数据带入回归方程计算 2013 年 11 月 8 日的收盘价,再与真实的收盘价进行比较. 可以发现随着模型的不断改进残差越来越小 DVTEMA 是 5 个模型中较优的一个.

4.2　多元线性回归模型

选取浦发银行作为样本股,用收盘价作为参考价格,用 2013 年 1 月 21 日至 2013 年 11 月 6 日的 188 个交易日的数据建立回归模型,用 2013 年 11 月 7 日、8 日、11 日的数据进行预测验证. 多元线性回归模型如下:

$$Y = 1.694 + 0.847X_1 - 2.272X_2 - 0.008X_3 + 0.003X_4, \quad (14)$$

其中, Y 为次日收盘价; X_1 为 DVTEMA-10; X_2 为 MACD; X_3 为 KDJ-D; X_4 为 RSI1(6 日相对强弱指标).

由于是用当日的股票数据去预测次日的涨跌,为了预测 2013 年 11 月 7 日、8 日、11 日的股票价格,就必须使用 6 日、7 日、8 日的数据. 预测结果:

表 2　浦发银行的预测情况

日期	预测值	真实值	相对误差
2013/11/7	10.113 61	10.02	0.93%
2013/11/8	10.195 64	9.98	2.16%
2013/11/11	8.741 753	10.07	13.19%

随着预测周期的增加,相对误差也逐渐增加. 整个模型的出发点是预测次日收盘价,误差会随着预测周期的增长而加大. 而模型中使用的短期指标居多,如进行中长期的预测,需将模型涉及的短期指标全部换为中长期指标.

4.3　多元非线性回归模型

构造回归模型的过程中需要用到 SPSS 的辅助. 使用浦发银行 2013 年 1 月 21 日至 2013 年 11 月 5 日的 188 个交易日的数据. 次日收盘价关于 DVTEMA 的曲线拟合 S 模型:

$$Y_1 = e^{3.157-\frac{8.569}{X_1}}. \quad (15)$$

同理,对 X_2、X_3、X_4 进行相同的操作,得到如下三个曲线模型:

次日收盘价关于 MACD 的曲线拟合 Cubic 模型:

$$Y_2 = 9.659 - 5.616 \times X_2 + 0.489 \times X_2^2 + 13.916 \times X_2^3. \quad (16)$$

次日收盘价关于 KDJ-D 的曲线拟合 Cubic 模型:

$$Y_3 = 8.969 + 0.129 \times X_3 - 0.004 \times X_3^2 + 0.000\,025\,4 \times X_3^3. \quad (17)$$

次日收盘价关于 RSI1 的曲线拟合 Cubic 模型:

$$Y_4 = 10.134 + 0.026 \times X_4 - 0.001 \times X_4^2 + 0.000\,006\,93 \times X_4^3. \quad (18)$$

用 Y 对 X_1、X_2、X_3、X_4、Y_1、Y_2、Y_3、Y_4 做线性回归,回归的方法是使用逐步回归法,逐步回归[6]最终选中的变量为 Y_1、MACD、Y_3 和 DVTEMA,R(描述数据拟合程度)达到了 0.964,比之前的线性模型提高了 0.004. 最终的多元线性回归模型为:

$$Y = -1.096 + 2.250Y_1 - 2.370X_2 + 0.086Y_3 - 1.207X_1. \quad (19)$$

其中 $Y_1 = e^{3.157-\frac{8.569}{X_1}}$; $Y_3 = 8.969 + 0.129 \times X_3 - 0.004 \times X_3^2 + 0.000\,025\,4 \times X_3^3$. 将 Y_1 和 Y_3 带入(19)式得:

$$Y = -0.324\,666 + 2.250e^{3.157-\frac{8.569}{X_1}} - 2.370X_2 + (0.011\,094 \times X_3 - 0.000\,344 \times X_3^2 + 0.000\,002\,184\,4 \times X_3^3) - 1.207X_1. \quad (20)$$

该模型在预测次日的股价时误差较小,当预测周期

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

稍微增加就会出现较大的误差,是因为 Y_1 中有指数项,容易放大误差. 可以发现与线性回归模型相比,虽然模型的精确性由 0.960 提高到了 0.964,但在实际预测次日的收盘价时差异不大,随着预测周期的增加,反倒是线性模型的准确性比非线性模型的准确性高.

5 结束语

随着经济全球化和数字设备的进步,金融数据的积累达到了前所未有的速率. 快速增长的数据量远远超过了人类分析它们的能力. 金融时间序列数据比其他统计数据更复杂,因为金融数据的变化更为不规则.

改进后的移动平均线在短期预测方面具有一定优势,在考虑对所持股票的买入和卖出时,依据的仅仅是股价简单的涨跌,然而现实中的股市行情纷繁复杂,简单的涨跌信号是不足以作为充分的判断依据的. 还应该考虑股票涨跌的趋势. 所谓趋势可以简单的理解为三种情况: 上涨的趋势,下跌的趋势,没有明显上涨下跌趋势[7]. 最理想的买入点并不一定是股价刚刚下跌的时候,持股者往往希望在连续下跌形成一定的下跌趋势,当跌至谷底即将反弹的时候才是一个较佳的买入时机; 反之,对于卖出点也是同样的道理.

参 考 文 献

[1] 孙碧波. 移动平均线有用吗? ——基于上证指数的实证研究[J]. 数量经济技术经济研究, 2005, 2: 149 – 156.

[2] 邹杨,王未卿,马琳. 对移动平均线的改进[J]. 中国证券期货,2011,5: 187 – 188.

[3] 张洪春. 实证指数平滑移动平均线在中国股市的有效性[D]. 上海: 华东理工大学,2012: 3 – 4.

[4] 周慧超. 基于移动平均线的趋势跟踪交易策略研究[D]. 上海: 上海交通大学,2014: 13 – 14.

[5] 徐奔. 基于 SPSS 相关性和回归分析的股票投资算法研究[J]. 电脑知识与技术,2015,18: 88 – 89.

[6] 江雪娇,孙涛,吴启. 基于逐步回归分析法的钢铁业股票价格研究[J]. 价格月刊,2010,2: 24 – 26.

[7] Dash R, Dash P K. A hybrid stock trading framework integrating technical analysis with machine learning techniques[J]. The Journal of Finance and Data Science,2016,3: 9 – 16.

A Multivariate Nonlinear Regression Model Based on the Improvement of Moving Average

Li Xinyi Zhou Qing

(School of Science, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract

Moving average is commonly used in technical analysis to predict the tendency of stock price. But ordinary moving averages only consider the relationship between closing price and time. Because of the limitations, we put forward the improved of moving averages step by step. The first step is to make a weighting for Moving Average by using stock turnover to get VTMA. The second step is to use the calculating formula of Exponential Moving Average to improve the formula in the first step to get VTEMA. The third step is start from the computation schema of triangle moving average to make a double smoothing for VTEMA in the second step to get DVTEMA. In order to predict the stock's closing price of the next day, multivariate nonlinear regression model is established based on DVTEMA, MACD, KDJ and RSI.

Key words: moving average, turnover, exponential moving average, multivariate nonlinear regression model.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net