

# 基于优化 MACD 模型的股票价格变化趋势预测方法\* Predicting Stock Price Trend by Optimizing MACD Model

杨永洁

YANG Yongjie

(广西大学计算机与电子信息学院, 广西南宁 530004)

(School of Computer, Electronics and Information in Guangxi University, Nanning, Guangxi, 530004, China)

**摘要:**【目的】探寻一种有效预测股票价格变化趋势的方法。【方法】在指数平滑异同移动平均线(MACD)指标中加入市场活跃程度(ACT)、波动率(VOL)、离差值(DIF)趋势程度3个指标来构建股票价格变化趋势预测模型。采用热点图对MACD策略中的参数以及股票进行选择、融合,并运用技术分析工具、支持向量机(SVM)与相关向量机(RVM)等机器学习方法对MACD策略中产生的交易信号进行优化,筛选特征变量。【结果】将设计的股票价格变化趋势预测策略的数据换成A股全市场数据并进行回测,发现近10年的年化收益率(14.8%)胜过沪深300指数(7.201%),而且使用A股全市场的数据有效避免了幸存者的偏差。【结论】优化改进的股票价格变化趋势预测模型可以在一定程度上预测股票上升的趋势,有效规避风险。

**关键词:** 股票价格 量化交易 机器学习

**中图分类号:** TP181 **文献标识码:** A **文章编号:** 1002-7378(2017)01-0065-06

**Abstract:**【Objective】Explore a way to effectively predict stock price trends.【Methods】Three indices of market activity, volatility, and DIF trend are added into the Moving Average Convergence and Divergence (MACD) strategy to generate stock price trend prediction model. The parameters of MACD and a subset of stocks are selected by the heat map. Technical analysis and machine learning methods such as support vector machine and relevance vector machine are used to improve the signaling capability of the MACD strategies.【Results】We obtain 14.8% annualized out of sample back analysis return in Chinese A-share market for the period of 1/4/2006 to 8/31/2016, which is significantly higher than Hushen300 annual return of 7.201% for the same period. Survivorship bias is avoided with all stocks in Chinese A-share market.【Conclusion】Our results demonstrate that to a certain extent the improved stock price trend prediction model can capture the upward tendency of Chinese stock market and effectively reduce risks.

**Key words:** stock price trending, quantitative trading, machine learning

收稿日期: 2017-01-10

作者简介: 杨永洁(1994-)女, 硕士研究生, 主要从事网络与并行计算、大数据分析研究, E-mail: 805093932@qq.com。

\* 广西科学研究与技术开发计划项目(桂科能 1140008-3B)资助。

## 0 引言

【研究意义】股票市场具有很高的风险,若股票价格波动太大或稳定性太低,它的风险会变得不可控。全球大多数股票都具有惯性效应,它是价格趋势的一种特殊情况。这些金融资产收益率在短期(1

年以内)内具有持续性,而从长期(1年以上)看会出现反转<sup>[1]</sup>。从 Jegadeesh<sup>[2]</sup>后,许多研究者开始从不同的方面对金融资产的可预测性进行探索。【前人研究进展】Hong 等<sup>[3]</sup>试图用一种统一的理论解释股票市场中的过度反应、动量交易以及反应不足的情况。Asness 等<sup>[4]</sup>指出:股票、外汇、大宗商品、股指期货、股指以及债券中均存在趋势。姜富伟等<sup>[5]</sup>研究了中国股票市场中样本内和样本外的可预测性。韩豫峰等<sup>[6]</sup>通过研究中国股市的股票特征来判定惯性与特征的相关性,发现惯性与这些特征极为相关,指出中国股票市场与其他国家股票市场都存在趋势。但该文并没有对如何预测中国股票市场的趋势(或短期惯性)进行研究。【本研究切入点】通过构建若干技术指标,改进优化指数平滑异同移动平均线(MACD)模型,融合运用线性回归和机器学习方法,对中国股票市场价格变化趋势进行预测。【拟解决的关键问题】基于 MACD 模型,引入新增交易活跃程度(ACT)、波动率(VOL)和离差值(DIF)趋势程度3个指标来控制产生股票交易的信号,并利用机器学习方法对显著影响惯性效应的特征(股票特征、股票投资者的类型、宏观经济政策)进行筛选,再对 MACD 策略中产生的买卖点进行改进优化,以使其收益更高且(或)风险更低,提出一种股票价格趋势预测方法。

## 1 优化的 MACD 模型

### 1.1 数据预处理

为了使股票的波动趋势更具多样性,设计的交易策略更具有普遍性,挑选43个行业(电力、房地产、钢铁、港口业、化工、酿酒、有色金属、银行等)共93种股票,2005年1月到2016年1月的分钟级数据。所有数据均为上海证券交易所和深圳证券交易所的上市股票,且从 Wind 资讯平台下载(该平台发布信息和数据准确率达到99.95%)。

由于某些股票分钟级市场活跃度低、交易频率被限制,所以将提取的93只分钟级样本数据处理成日线级别,即基于日线来设计交易策略。获取数据后,将数据整理成标准输入矩阵,前期还需要将股票按照全时间序列排列,并对数据进行补洞处理。

### 1.2 优化 MACD 指标

通过 Excel 的 VBA 编程对 MACD 指标优化并建模。MACD 指标由离差值(DIF)、异同平均数(DEA)、MACD 柱状图,两线一柱3部分组成。 $EMA_{M_t}$  和  $EMA_{N_t}$  分别为第  $t$  天的  $M$  日指数移动平

均以及  $N$  日指数移动平均,其中  $M > N$ 。

$$EMA_{N_t} = EMA_{N_{t-1}}(N-1)/(N+1) + 2C_t/(N+1), \quad (1)$$

$$EMA_{M_t} = EMA_{M_{t-1}}(M-1)/(M+1) + 2C_t/(M+1), \quad (2)$$

$$DIF_t = EMA_{N_t} - EMA_{M_t}, \quad (3)$$

$$DEA_{P_t} = DEA_{P_{t-1}}(P-1)/(P+1) + 2DIF_t/(P+1), \quad (4)$$

$$MACD_t = 2(DIF_t - DEA_{P_t}). \quad (5)$$

当短周期的 EMA 穿过较长周期的 EMA 时,会出现交叉信号。交叉信号的方向由移动平均交叉的方向决定。当较短周期的 EMA 线向上穿越较长周期的 EMA 线时,会产生积极交易信号,即买入信号,意味着出现逐渐上升的积极时机;当较短周期的 EMA 线向下穿越较长周期的 EMA 线时,会产生消极信号,即卖出信号,意味着出现逐渐上升的消极时机。

基于 MACD 指标的股票价格趋势预测策略中,涉及到两种收盘价格,一种用于产生交易信号的收盘价格,另一种作为计算股票交易累积收益时的收盘价格。原因在于加载策略到买入股票的过程中,会产生时间差,导致产生交易信号的价格与最终交易买入股票的价格有差别。本研究采用两种收盘价格来建立交易模型,使 MACD 策略算出的累积收益更加接近真实交易时的收益。

由于 MACD 是一种具有明显趋势的策略,当行情为趋势型时有很好的效果,但当行情为震荡型时会产生很多假信号。因此引入股票市场活跃程度(ACT)、波动率(VOL)、DIF 趋势程度( $K$ )指标来辅助 MACD 指标,使产生更加稳健的交易信号。又由于交易者比较倾向于活跃度高的市场,故取某一时刻的数据来计算市场活跃度具有突发性和偶然性,并不能代表这段时间市场的具体情况。设  $H_{M_t}$  和  $L_{N_t}$  分别为第  $t$  天股票价格最高价的  $M$  日指数移动平均,第  $t$  天股票价格最低价的  $N$  日指数移动平均,  $StdC_{P_t}$  为第  $t$  天收盘价标准差的  $P$  日指数移动平均:

$$ACT = (H_{M_t} - L_{N_t}) / StdC_{P_t}. \quad (6)$$

若股票的活跃性比较大,波动性也会变大。此时若缺少风险管理工具,就会存在较大的风险,使市场失去吸引力。股票价格的波动越剧烈,波动率就越高,资产收益率的不确定性就越强。本研究主要研究历史波动率,即根据时间序列来分析处理历史数据,并用 22 d 收盘价移动平均除以当天的收盘价来计算

历史波动率。 $C_{Q_t}$  为第  $t$  天股票价格收盘价的  $Q$  日指数移动平均:

$$VOL = C_{Q_t} / C_t. \quad (7)$$

DIF 是快均线减去慢均线的差,当股票价格的整体趋势趋于平稳,但波动比较大时,快均线和慢均线会出现短时间内反复交叉的情况。当行情震荡波动严重时,会产生大量的假信号导致交易成本急剧上升。为此,增加一个 DIF 趋势程度指标,用于度量 DIF 的斜率,当斜率比较大时,即其当与  $0$  轴所围成的夹角比较大时,认为此时产生的交易信号是可信的。使用  $22 d$  收盘价的斜率除以当天的收盘价来计算 DIF 趋势程度。 $S_{R_t}$  为第  $t$  天股票价格收盘价的  $R$  日指数移动平均的斜率:

$$K = S_{R_t} / C_t. \quad (8)$$

## 2 参数选择与预测策略

对交易策略的参数以及股票进行挑选,并对挑选出来的股票所产生的交易信号以及收益进行记录。由于并非每一个交易信号产生的收益均为正,所以通过采取机器学习方式对产生的交易信号进行预测。若预测的收益为负,且 MACD 策略产生的收益为负,则去除该交易点。

### 2.1 参数选择

交易之前剔除风险大于回报的股票,缩小选择股票的范围。基于统计分析方法,采用 VBA 编程对 Excel 数据表中股票进行筛选。

MACD 策略有 3 个参数,分别为 Long, Middle 和 Short。为了得到最佳的收益,对于不同特征的股票数据,这 3 个参数值的选择也不尽相同。比如 MACD(5, 35, 5) 比 MACD(12, 26, 9) 具有更高的敏感性,更加适合于周线。3 个参数的变化范围见表 1。又使用变量 Pnl 来表示累积收益,其中累计收益 Pnl 是关于 Long、Middle、Short 的函数,记为 Pnl(Long, Middle, Short)。

对于每一种股票,都需要加载 MACD 策略,并用三重循环实现参数(Long, Middle, Short)的变化,每一组参数都对应一个 Pnl(Long, Middle, Short) 值,最终算出该只股票所有的 Pnl 值。采用热点图方式选出最好的 Pnl 值,并记录下它所对应的(Long, Middle, Short)的值。为方便表示热点图,需要固定 1 个点,将 3 维变成 2 维。选择了 5 个 Middle 的值和 5 个 Short 的值,如表 2 所示。

图 1 中的每个小方格均代表一个 Pnl(Long, Middle, Short) 值。参数 Middle 固定为 42, Short、

Large 的值在规定的范围内变动,从图 1 中筛选出 3 对(Long, Short)的值。同理固定参数 Short,从对应的热点图中选出 3 对(Long, Middle)。热点图的颜色变化表示收益变化,绿色表示累积收益较高,即 Pnl(Long, Middle, Short) 较大的参数对,红色表示累积收益较低,即 Pnl(Long, Middle, Short) 较小的参数对。分别固定选出的 5 个 Short 的值,在参数变化范围内变动(Long, Middle)的值。同理,分别固定选出的 5 个 Middle 的值,在参数变化范围内变动(Long, Short)的值。最终做出 10 个热点图,从每个热点图中选出 3 对(Long, Middle)以及 3 对(Long, Middle),最终选出 30 对(Long, Middle, Short)值,见表 3。

表 1 (L, M, S) 变化范围

Table 1 Range of (L, M, S) value

参数	最小值	最大值
Parameters	Min	Max
Long	10	80
Middle	3	78
Short	3	78

表 2 Short 和 Middle 参数值

Table 2 Values of short and middle

Short	Middle
7	9
11	18
19	42
37	60
51	75

### 2.2 股票选择

从 93 种股票数据的前 100 个 Pnl(L, M, S) 中,找到与表 3 中(L, M, S) 匹配的参数值所对应的年化收益率,即 Pnl 值,并计算它们的最大值(Max)、最小值(Min)、平均值(Avg)。对应计算出的最大值、最小值、平均值排序就可以从 93 只股票集中挑选出相对表现好的股票或者适应性较好的股票,如表 4 所示。

### 2.3 寻找收益与特征变量的线性关系

在 R 语言中可以直接通过 lm() 函数探究自变量(特征变量) X 与因变量(累积收益) Y 之间的线性关系。本文的研究中包含 71 个自变量 X。采用多元回归方法分析他们之间的关系。图 2 给出了线性回归预测的因变量值 Y1 和真实值 Y 之间的关系。横轴表示预测的收益值 Y1、纵轴表示 MACD 模型

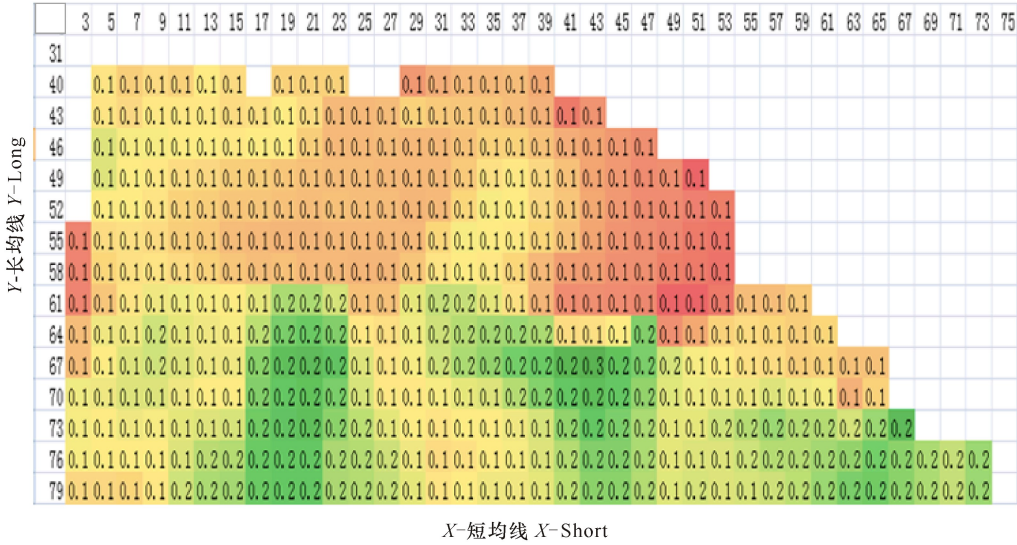


图 1 热点图示例

Fig. 1 Example of heap map

表 3 选择的参数值

Table 3 Selected parameter values

Short value is fixed			Middle value is fixed		
Short	Middle	Long	Short	Middle	Long
7	36	67	37	9	76
7	45	49	47	9	76
7	48	70	61	9	79
11	33	70	31	18	76
11	48	76	41	18	73
11	63	76	63	18	76
19	39	76	21	42	67
19	45	67	65	42	76
19	57	67	43	42	70
37	42	67	43	60	64
37	21	76	53	60	76
37	66	70	33	60	76
51	12	76	29	75	79
51	63	67	51	75	79
51	69	76	67	75	79

表 4 挑选的股票代码

Table 4 Selected stocks ID

序号 No.	股票代码 Stocks ID	序号 No.	股票代码 Stocks ID
1	SH600012	15	SZ000063
2	SH600020	16	SZ000563
3	SH600055	17	SZ000792
4	SH600176	18	SZ000831
5	SH600177	19	SZ000858
6	SH600267	20	SZ000876
7	SH600487	21	SZ000983
8	SH600550	22	SZ002018
9	SH600489	23	SZ002021
10	SH600519	24	SZ002024
11	SH600550	25	SZ002029
12	SH600808	26	SZ002049
13	SH600970	27	SZ002216
14	SH600824		

产生的收益 Y。若 Y1 与 Y 的值完全相同,那么画出的图形就应该是一条离散的直线。从图 2 可以判断,若 MACD 模型产生的收益 Y 为负值,且预测的收益值 Y1 也为负值,那么该交易信号将被取消,即以 0 点为中心点的第三象限的交易信号将被全部取消。

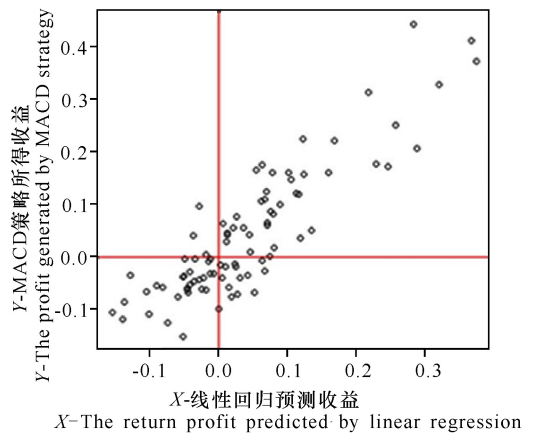


图 2 线性回归预测

Fig. 2 Linear regression prediction

### 2.4 寻找收益与特征变量的非线性关系

支持向量机(SVM)是一个有监督作用的学习模型,它既可以分析线性情况,也可以对线性不可分的情况进行分析,可以获得二值输出。相关向量机(RVM)是一种与 SVM 类似的监督学习方法。RVM 与 SVM 的区别:RVM 在贝叶斯框架下进行训练,它可以获得二值输出以及概率输出。

由于影响股票数据的因素之间大多存在非线性关系,为此,采用 RVM 和 SVM 方法寻找出累积收益与特征变量之间的非线性关系。图 3 和图 4 分别给出使用 RVM 和 SVM 对因变量 Y 进行预测的结果。

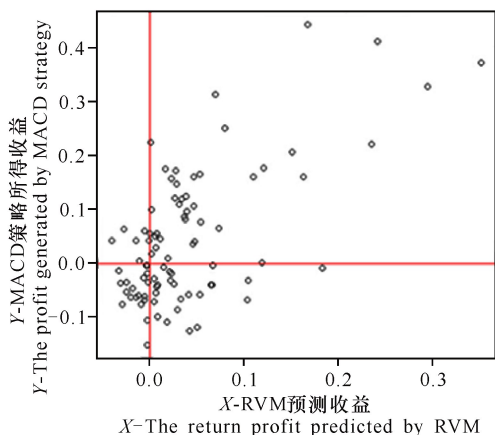


图 3 利用 RVM 预测的结果

Fig. 3 Prediction results by RVM

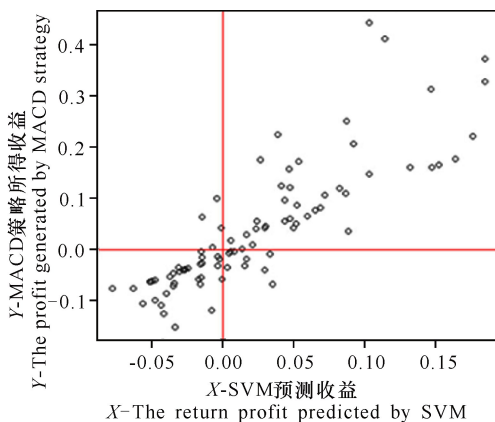


图 4 利用 SVM 预测的结果

Fig. 4 Prediction results by SVM

由图 3 和图 4 可以看出,SVM 比 RVM 的预测能力更强,其中 SVM 的点可以用一元一次函数拟合,预测的 Y1 值比较接近真实的 Y 值。同理,从图 3、图 4 中可以判断,若 MACD 模型产生的收益 Y 为负值,且通过 SVM 或者 RVM 预测的收益值 Y1 也为负值,那么该交易信号将被取消,即以 0 点为中心点的第三象限的交易信号将被全部取消。

### 3 验证分析

机器学习训练模型的整个过程,通过 R 语言在 R-Studio 平台上实现。用线性回归、SVM 或者 RVM 对交易信号进行改进。若 MACD 模型产生的收益 Y 为负值,且通过线性回归、SVM 或者 RVM 预测的收益值 Y1 也分别为负值,那么通过线性回归、SVM 或者 RVM 进行预测后为负的交易信号将被取消,即以 0 点为中心点的第三象限的交易信号将被全部取消。最后将线性回归、RVM、SVM 各自对 MACD 策略的交易信号进行优化后得到的累积收益与 MACD 策略的累积收益进行比较。收益比较结果见图 5。

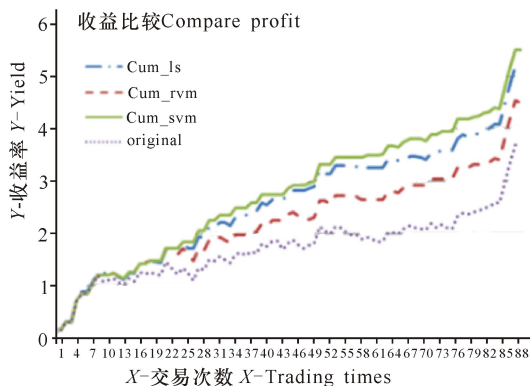


图 5 累积收益比较

Fig. 5 Comparison for market accumulated earning

图 5 中曲线 Cum\_ls 表示采用线性回归方法对 MACD 策略交易产生的交易信号进行改进后得到的累积收益,Cum\_rvm 表示采用 RVM 对 MACD 策略交易产生的交易信号进行改进后得到的累积收益,Cum\_svm 表示采用 SVM 对 MACD 策略交易产生的交易信号进行改进后得到的累积收益,original 表示采用 MACD 策略交易产生的交易信号的累积收益。从图 5 可看出,使用机器学习方法对 MACD 策略进行改进后的收益均比原模型有一定的改善;另一方面,采用线性回归、RVM 方法改进的效果相对较小,而采用 SVM 方法改进的效果较大。

图 6 给出了代码为 002025 股票价格的曲线。将图 5 累积收益与图 6 的股票价格在时间上进行对比,可以看出:当股票价格呈上升趋势时,收益增幅较大。例如,2015 年 7 月到 2016 年 9 月股票价格持续上涨明显,收益也大幅增加。说明改进的交易策略能够有效避开价格震荡区域,如 2007 年 2 月到 2007 年 12 月的区域;在区间 2008 年 5 月到 2008

年8月股票价格下跌时,采用 SVM 改进模型获得的交易信号能够有效的避免下跌风险。

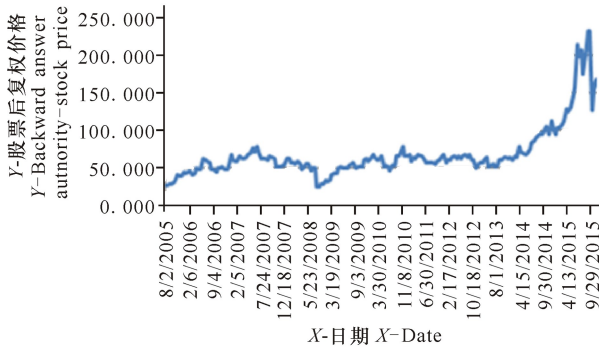


图6 股票价格

Fig. 6 Stock Price

## 4 结论

虽然股票市场是一个复杂的非线性动态系统,但可以通过抓住股票市场的趋势来提高股票投资的准确性。本文通过加入3个指标来控制产生的交易信号,改进了交易策略,通过画出累积收益的热点图对参数 Long、Middle 和 Short 以及股票进行选择;运用线性回归、RVM、SVM 方法对 MACD 策略的交易信号进行优化,从而使交易模型预测获得的收益更高且(或)风险更低。当数据中噪音数量过多时,SVM 的方差会受噪音的影响而变大,结果也会受到影响,这是后续工作需要研究解决的问题。此外,本文给出的交易策略是基于历史数据建模,下一步将研究通过不断更新数据和训练模型,以获得更

好的预测效果。

### 参考文献:

- [1] MOSKOWITZ T J, OOI Y H, PEDERSEN L H. Time series momentum[J]. *Journal of Financial Economics*, 2012, 104(2): 228-250.
- [2] JEGADEESH N. Evidence of predictable behavior of security returns[J]. *The Journal of Finance*, 1990, 45(3): 881-898.
- [3] HONG H, STEIN J C. A unified theory of underreaction, momentum trading, and overreaction in asset markets[J]. *The Journal of Finance*, 1999, 54(6): 2143-2184.
- [4] ASNESS C S, MOSKOWITZ T J, PEDERSEN L H. Value and momentum everywhere[J]. *The Journal of Finance*, 2013, 68(3): 929-985.
- [5] 姜富伟, RAPACH D E, STRAUSS J K, 等. 中国股票市场可预测性的实证研究[J]. *金融研究*, 2011(9): 107-121.  
JIANG F W, RAPACH D E, STRAUSS J K, et al. How predictable is the Chinese stock market? [J]. *Journal of Financial Research*, 2011(9): 107-121.
- [6] 韩豫峰, 汪雄剑, 周国富, 等. 中国股票市场是否存在趋势? [J]. *金融研究*, 2014(3): 152-163.  
HAN Y F, WANG X J, ZHOU G F, et al. Are there trends in Chinese stock market? [J]. *Journal of Financial Research*, 2014(3): 152-163.

(责任编辑:尹 闯)