

Do the Analysts' Recommendations Really Mean Something?

—An Empirical Study from the Chinese Stock Market

Jiaping Zhang

School of Mathematics and Statistics, Hainan Normal University, Haikou
Email: zhjp1111@yeah.net

Received: Apr. 14th, 2014; revised: May 20th, 2014; accepted: Jun. 1st, 2014

Copyright © 2014 by author and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we randomly selected 30 stocks from the Chinese securities market. We carried out research about the relation between the stock yield rate and the investment rate in the laws of the statistics, using its investment rate reports and stock prices from December 25th in 2012 to December 25th in 2013. Among them, in order to ensure the validity of the statistical analysis, we use the famous Box-Cox Transformation in the statistical analysis and adopt the optimization method. By maximizing the F-statistic, we could choose the optimal linear regression model among the stock yield rate, the investment rate and the previous stock yield. The empirical analysis results show the correlation between stock returns and its early returns; however, the analysts' investment rate has nothing to do with the subsequent performance of the stock yield rate.

Keywords

Yield Rate, Investment Rate, The Box-Cox Transformation, Linear Regression Model, F-Statistic

金融股票分析师的建议真的有投资价值吗？

—来自中国证券市场的一个实证报告

张佳平

海南师范大学数学与统计学院，海口
Email: zhjp1111@yeah.net

收稿日期：2014年4月14日；修回日期：2014年5月20日；录用日期：2014年6月1日

摘要

本文随机选取了中国证券市场30支¹股票，就其2012年12月25日至2013年12月25日的投资评级报告和股票价格数据来对股票的收益率与投资评级间的关系进行了统计规律的研究。其中，为保证统计分析的有效性，采用了统计分析中著名的Box-Cox变换，并使用了最优化方法。通过极大化F统计量来构造股票收益率与投资评级、前一期收益率间的最优线性回归模型。实证分析结果表明，股票收益率的后续表现与相应前一期的收益率相关，而与分析师的投资评级无关。

关键词

收益率，投资评级，Box-Cox变换，线性回归模型，F统计量

1. 引言

华尔街有这么一句家喻户晓的谚语：不要依靠金融股票分析师的投资评级来买卖股票，但是也不要抛弃它们。股票投资评级中常常包含着许多重要的信息，而投资评级信息的准确性，即其可信度需要第三方给出客观地评判[1]。

股票市场是信息高度密集型的虚拟市场，股价的涨落对于信息的波动是十分灵敏的，而作为向投资者提供上市公司信息的群体，股票分析师在股票市场中扮演着重要的角色。在美国，金融市场发展日趋成熟，分析师群体对股票市场有很大的影响力。随着中国证券股票市场的发展和成熟，越来越多的投资者(特别是机构投资者)已经开始积极关注并利用分析师的投资报告。但是，许多投资者在获得投资评级数据后，常常因为不能正确利用这些数据，反而会对投资造成不利影响。分析师的投资评级都是相对于一定的时间段给出的，有一个月、三个月、六个月和十二个月，不同投资风格的投资者选择的持股时间是不相同的。即便是在金融市场最发达的欧美，评级为买入的股票也并不一定就比评级为卖出的股票的后续市场表现得更好[2]。

评估这些投资评级报告与市场实际走向吻合的程度是本文的研究目标。本文通过随机选取中国证券市场30支股票，就其2012年12月25日至2013年12月25日间一个月、三个月、六个月、十二个月的投资评级报告和股票价格数据来对股票收益率和投资评级间的关系进行统计规律的研究，基于这些实证数据分析，判断金融股票分析师的投资评级建议是否有投资价值。

2. 数据准备[3]

本研究的数据准备工作主要包括以下步骤：

步骤一：样本选择

随机选择了中国证券市场中共计30家上市公司的股票，每一家上市公司建立一个独立的报告库，收录相关的股票价格数据和投资评级报告，时间从2012年12月25日至2013年12月25日。

步骤二：标准化

对每份报告设置以下五个要素来描述：报告来源、报告日期、股票名称、股票价格、投资评级。

¹ 随机选取的30支股票取自大智慧365软件，其代码为：600000、600004、600016、600028、600029、600031、600036、600048、600050、600060、600064、600071、600085、600104、600109、600111、600119、600127、600130、600150、600159、600162、600176、600183、600192、600198、600221、600261、600345、600401。

步骤三：数据量化

每份股票评级报告中的分析师的不同统计时段(一个月内、三个月内、六个月内、十二个月内)的评级有五个等级“买入”、“增持”、“中性”、“减持”、“卖出”分别对应给予赋值“-2”、“-1”、“0”、“1”、“2”。这样量化可以清晰地表明股票投资评级等级间的差别，便于理解。

步骤四：整合数据

数据量化完成后，得到量化后的投资评级报告。对同一支股票，收集如下数据：(1) 评级分，计算不同统计时段投资评级的平均得分；(2) 股票价格数据，计算固定时段的收益率。

步骤五：指标化[4]

不同股票、不同统计时段的投资评级的平均得分记作 x_{ij} ，

不同股票、不同统计时段的股票收益率记作 z_{ij} ，记平移后的股票收益率为 y_{ij} ， $y_{ij} = z_{ij} + 1$ ，

不同股票、不同时段所对应的前一期的股票收益率记作 r_{ij} 。

其中 $i = 1, 2, \dots, 30$ ， $j = 1, 3, 6, 12$ 。

3. Box-Cox 变换[5]

通常，在普通的回归模型中，往往使用如下的线性模型：

$$Y = \beta X + e \quad (1)$$

给定观测得到的试验数据集 $(x'_i, y_i), i = 1, 2, \dots, n$ ，如果用上面的模型(1)来进行拟合，那么就要求给定的数据满足下列条件：

(a) 线性相依性，即对(1)式两端同时取期望，得到 $E(Y) = \beta E(X)$ ；

(b) 等方差性，即要求误差项 e 的各分量是等方差且相互独立；

(c) 正态性，要求误差项 $e \sim N(0, \sigma^2 I)$ 。

但是在处理实际问题时，对观测得到的试验数据集 $(x'_i, y_i), i = 1, 2, \dots, n$ ，经过回归诊断后得知，某些数据不满足 Gauss-Markov 条件，需要对数据采取“治疗”措施，即通过检验发现，观测获取的某些数据不完全满足上述线性相依性、等方差性和正态性的条件。实践证明，数据变换是处理有问题数据的一种好方法。出于有效的提高回归精度的目的，可以采取一种常用的数据变换的工具：经典的 Box-Cox 变换。该变换的主要特点是引入一个新的参数 λ ，通过数据本身估计出参数 λ ，继而确定应该采取的数据变换形式。实践证明，Box-Cox 变换对许多实际数据都是行之有效的：对因变量的数据变换能显著地改善数据的正态性、方差齐性以及对称性；同样，对自变量的数据变换能够改善模型的结构，使拟合效果变好。

定义 1：设 $Y = (y_1, \dots, y_n)$ ， $y_i > 0$ ， $i = 1, \dots, n$ ，变换

$$Y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln Y, & \lambda = 0, \end{cases} \quad (2)$$

此处的 λ 为一个待定变换参数。称式(2)为 Y 的 Box-Cox 变换(也称幂变换)。

Box-Cox 变换是一族变换，包括很多常见的变换，譬如对数变换($\lambda = 0$)，倒数变换($\lambda = -1$)和平方根变换($\lambda = 1/2$)。

通过因变量 Y 的 Box-Cox 变换后， $Y^{(\lambda)}$ 满足下面的线性模型：

$$Y^{(\lambda)} = \beta X + e, \quad e \sim N(0, \sigma^2 I) \quad (3)$$

其中 λ 为变换参数，且 $-2 \leq \lambda \leq 2$ ，Box-Cox 变换的主要目的是找出合适的的数据变换参数 λ 的估计。由于 $Y^{(\lambda)} \sim N(\beta X, \sigma^2 I)$ ，因此可以用极大似然(MLE)的方法来确定 λ 。继而使模型(3)的回归拟合效果变得更

好。

注意(2)式中的 Box-Cox 变换仅适用于正值变量，对于变量可以取负值的情况，往往使用带有平移参数的变换取代(2)，即给变量 y_i 加上平移参数 c ，使得 $y_i + c > 0$ ，此时(2)式变为

$$Y^{(\lambda)} = \begin{cases} \frac{(Y+c)^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln(Y+c), & \lambda = 0, \end{cases} \quad (4)$$

平移参数 c 使得任意地 $y_i + c > 0$ ， $i = 1, \dots, n$ ，即要求 $c > -\min(y_i)$ 。

以上是仅对因变量 Y 进行 Box-Cox 变换，有时也称为单边变换。有时基于实际问题的数据及回归拟合效果的考虑，也可对自变量 X 进行 Box-Cox 变换，即

$$X^{(\phi)} = \begin{cases} \frac{X^\phi - 1}{\phi}, & \phi \neq 0, \\ \ln X, & \phi = 0, \end{cases} \quad (5)$$

此时，模型(3)变为双边 Box-Cox 变换的广义线性模型

$$Y^{(\lambda)} = \beta X^{(\phi)} + e \quad (6)$$

双边参数 ϕ 、 λ 的估计方法与单边变换的估计方法一致。

4. F 统计量[6]

在本文的检验评判标准中，将使用一个重要的回归统计量来衡量模型的优良性，即 F 统计量。

F 统计量函数

$$F = \frac{SS_{reg}}{\hat{\sigma}^2} = \frac{SYY - RSS}{\frac{RSS}{n-2}} = (n-2) \left(\frac{SYY}{RSS} - 1 \right) \quad (7)$$

其中：

$SS_{reg} = SYY - RSS$ ：表示回归平方和；

SYY ：表示 y_i 的校正平方和，且 $SYY = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$ ；

RSS ：表示回归残差平方和。

在线性回归模型(1)中，按照一般的显著性检验的步骤，在事先给定的显著性水平 α 下，当 $F > F_{1-\alpha}(p, n-p-1)$ 时，就认为 y 和 x_1, \dots, x_p 之间确实存在线性关系[7]。

F 统计量函数是回归分析的重要指标，在实际的回归模型中，较大的 F 值，表明线性回归模型拟合的效果较好。所以，在本文的模型选择中，一个重要的依据就是使回归模型的 F 统计量函数达到极大值的 λ 作为因变量 y 的 Box-Cox 变换中的数据变换参数。从根本上来讲，本文中模型的选择相当于解最优化问题：

$$\max_{\lambda} F(\lambda)$$

得出 λ 。

5. 实证分析

选用不同时段对应的前一期股票收益率 r_{ij} 和投资评级的平均得分 x_{ij} 一同来对加上平移参数后的当期股票收益率 y_{ij} ($y_{ij} = z_{ij} + 1, y_{ij} > 0$) 做二元线性回归。模型如下：

$$y_{ij} = \alpha_0 + \alpha_1 r_{ij} + \alpha_2 x_{ij} + e_{ij}, \quad i = 1, 2, \dots, 30, j = 1, 3, 6, 12.$$

对于不同统计时段，判断模型的可行性，回归系数 α_1 、 α_2 的显著性，特别是投资评级变量的系数 α_2 的显著性。为方便说明，记 $p(\alpha_1)$ 为系数 α_1 的 p 值， $p(\alpha_2)$ 为系数 α_2 的 p 值。

不同统计时段的回归结果如下：

1) 当统计时段为一个月，即 $j=1$ 时，

$$y_{ij} = 2.0047 + 0.9987r_{ij} + 0.0006x_{ij} + e_{ij}, \quad R^2 = 1, F = 1683.7, p = 0.$$

通过对上述模型回归系数的显著性及置信区间分析可以得到： $p(\alpha_1) = 0 < 0.05$ ， $\alpha_1 = 0.9987 \in [0.963, 1.034]$ ，则回归系数 α_1 显著，从而表明一个月内，收益率 y_{i1} 与前一个月的收益率 r_{i1} 存在线性关系，并且两者间正相关；而 $p(\alpha_2) = 0.738 > 0.05$ ， $\alpha_2 = 0.0006 \in [-0.003, 0.004]$ ，则回归系数 α_2 不显著，从而表明一个月内，收益率 y_{i1} 与投资评级变量 x_{i1} 之间没有线性关系。

2) 当统计时段为三个月，即 $j=3$ 时，

$$y_{ij} = 1.9056 + 0.8978r_{ij} + 0.0032x_{ij} + e_{ij}, \\ R^2 = 0.9810, F = 698.6887, p = 0$$

通过对上述模型回归系数的显著性及置信区间分析可以得到： $p(\alpha_1) = 0 < 0.05$ ， $\alpha_1 = 0.8978 \in [0.848, 0.947]$ ，则回归系数 α_1 显著，从而表明三个月内，收益率 y_{i3} 与前三个月的收益率 r_{i3} 间存在线性关系，并且两者间正相关；而 $p(\alpha_2) = 0.555 > 0.05$ ， $\alpha_2 = 0.0032 \in [-0.008, 0.014]$ ，则回归系数 α_2 不显著，从而表明三个月内，收益率 y_{i3} 与投资评级变量 x_{i3} 之间没有线性关系。

3) 当统计时段为六个月，即 $j=6$ 时，

$$y_{ij} = 3.0551 + 2.1062r_{ij} + 0.0054x_{ij} + e_{ij}, \\ R^2 = 0.8257, F = 63.9707, p = 0$$

通过对上述模型回归系数的显著性及置信区间分析可以得到： $p(\alpha_1) = 0 < 0.05$ ， $\alpha_1 = 2.1062 \in [1.722, 2.491]$ ，则回归系数 α_1 显著，从而表明六个月内，收益率 y_{i6} 与前六个月的收益率 r_{i6} 间存在线性关系，并且两者间正相关；而 $p(\alpha_2) = 0.913 > 0.05$ ， $\alpha_2 = 0.0054 \in [-0.096, 1.106]$ ，则回归系数 α_2 不显著，从而表明六个月内，收益率 y_{i6} 与投资评级变量 x_{i6} 之间没有线性关系[8]。

4) 当统计时段为十二个月，即 $j=12$ 时，

$$y_{ij} = 1.1614 - 0.4540r_{ij} + 0.0321x_{ij} + e_{ij} \quad (8) \\ R^2 = 0.0806, F = 1.1843, p = 0.3214.$$

由拟合结果知， $p = 0.3214 > 0.05$ ，超出了通常可接受错误的边界水平，回归结果的可信程度减弱。并且通过对上述模型回归系数的显著性及置信区间分析可以得到： $p(\alpha_1) = 0.143 > 0.05$ ， $\alpha_1 = -0.4540 \in [-1.071, 0.163]$ ，则回归系数 α_1 不显著，从而表明十二个月内，收益率 y_{i12} 与前十二个月的收益率 r_{i12} 间没有线性相关关系；同样地 $p(\alpha_2) = 0.851 > 0.05$ ， $\alpha_2 = 0.0321 \in [-0.315, 0.379]$ ，则回归系数 α_2 也不显著，从而表明十二个月内，收益率 y_{i12} 与投资评级变量 x_{i12} 之间也没有线性关系。

对于统计时段为十二个月，即 $j=12$ 时，上述模型(8)的可信程度较弱，需要找寻更为恰当的模型。下面对(8)中因变量 y_{ij} 进行单边 Box-Cox 变换得到 $y_{ij}^{(\lambda)}$ ，其中变换参数 $-2 \leq \lambda \leq 2$ ，此时当统计时段为十二个月，即 $j=12$ 时，

$$y_{ij}^{(\lambda)} = 0.0152 - 0.4443r_{ij} + 0.0098x_{ij} + e_{ij} \quad (9) \\ R^2 = 0.2353, F = 4.1534, p = 0.0268, \lambda = -2, y_{ij}^{(\lambda)} = \frac{y_{ij}^{\lambda} - 1}{\lambda}.$$

此处的 $\lambda = -2$ 是通过 Matlab 编程，极大化 F 得到的。通过比较模型(8)、(9)的回归结果发现，对因变量收益率 y_{ij} 进行 Box-Cox 变换后，模型的各项评判指标的优良性均有提高，特别是 $p = 0.0268 < 0.05$ ，判定回归结果可信度高，具有真实的显著性和统计学意义[9]。

通过对上述模型(9)中回归系数的显著性及置信区间分析得到： $p(\alpha_1) = 0.008 < 0.05$ ， $\alpha_1 = -0.4443 \in [-0.763, -0.126]$ ，则回归系数 α_1 显著，从而表明十二个月内，Box-Cox 变换后的收益率 $y_{i12}^{(\lambda)}$ 与前十二个月的收益率 r_{i12} 之间存在线性关系，并且两者之间呈负相关关系；而 $p(\alpha_2) = 0.911 > 0.05$ ， $\alpha_2 = 0.0098 \in [-0.169, 0.189]$ ，则回归系数 α_2 不显著，从而表明十二个月内，Box-Cox 变换后的收益率 $y_{i12}^{(\lambda)}$ 与投资评级变量 x_{i12} 之间没有线性关系。

模型(8)，Box-Cox 变换前，不能确定十二个月内的收益率与前一期收益率之间是否存在线性关系，而在模型(9)中，Box-Cox 变换后，建模确定两者之间存在负相关的线性关系。由此可知看出，Box-Cox 变换对于建模寻找变量之间的相关关系，是一种可以尝试的有用途径。

6. 结论

基于本文中随机选取的 30 支股票 2012 年 12 月 25 日至 2013 年 12 月 25 日的实证分析结果表明，股票收益率与前一期收益率之间存在相关性，而分析师的投资评级与股票收益率的后续表现无关。基于此 [10]，投资者在作出投资决策时，不应该受分析师投资评级的干扰，应该根据股票市场可以得到的数据分析前期的股票收益率表现，作出合理的决策。短期中，一个月、三个月、六个月内，股票收益率与其相应的前期收益率之间都存在正相关关系，并且根据回归系数的区间估计得，统计时段为六个月时，股票收益率与前六个月的收益率之间的正相关性最强，投资者可以选择股票市场中短期收益率表现良好的股票进行投资或选取最优的投资组合；而长期中，十二个月内，股票收益率与其相应的前期收益率之间呈现负相关关系，表明股票的涨势一般不会持续太长时间，去年的明星股，一般不是今年的明星，投资者应更加关注短期内的明星股，继而作出优化、合理的投资决策。

参考文献 (References)

- [1] 丁亮, 孙慧 (2001) 中国股市股票推荐效应研究. *管理世界*, **5**, 111-116.
- [2] 金轶雯, 白峰杉 (2008) 金融证券分析师投资评级数据的分析挖掘. *数学的实践与认识*, **13**, 18-25.
- [3] 张焯, 胡倩, 周健 (2009) 证券分析师评级报告的投资价值研究——来自香港股市的经验数据. *当代财经*, **10**, 61-65.
- [4] 林翔 (2000) 对中国证券咨询机构预测的分析. *经济研究*, **2**, 56-65.
- [5] Cryer, J.D. and Chan, K.-S., 著, 潘红宇, 等, 译 (2011) 时间序列分析及应用. 机械工业出版社, 北京, 72-73.
- [6] 王松桂, 史建红, 尹素菊, 吴密霞 (2004) 线性模型引论. 科学出版社, 北京, 121-127, 151-154.
- [7] 陈希孺 (2009) 概率论与数理统计. 中国科学技术出版社, 合肥, 258-325.
- [8] Barber, B., Lehavy, R., McNichols, M. and Trueman, B. (2001) Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance*, **56**, 531-564.
- [9] Stickel, S.E. (1995) The anatomy of the performance of buy and sell recommendations. *Financial Analyst Journal*, **51**, 25-39.
- [10] Womack, K. (1996) Do brokerage analysts' recommendations have investment value? *Journal of Finance*, **51**, 137-167.