

The Analysis and Forecast of Bank Index and Its Constituent Stocks

Xiao Han, Xinghu Teng, Ting Dou

Army University of Engineering, Nanjing Jiangsu
Email: h.x.good@163.com

Received: Jul. 8th, 2020; accepted: Jul. 22nd, 2020; published: Jul. 29th, 2020

Abstract

Taking the k-line closing prices of bank indexes and constituent stocks as research samples, the bank index and its constituent stocks are analyzed. Based on the relationship between the bank index and its constituent stocks, a positive regression analysis was carried out, and the abnormal points in the sample data of the bank index and each constituent stock were screened out through regression diagnosis. In order to make the model more practical, ridge estimation is used to eliminate multicollinearity among variables, and the optimal subsets of variables are selected based on absolute constraint estimation and elastic constraint estimation. Finally, the linear regression equation between the bank index and its constituent stocks is given.

Keywords

Clustering Analysis, Regression Analysis, Ridge Estimate, Absolute Constraint Estimation, Elastic Constraint Estimation

基于银行类指数及其成分股的分析 and 预测

韩 笑, 滕兴虎, 窦 婷

陆军工程大学, 江苏 南京
Email: h.x.good@163.com

收稿日期: 2020年7月8日; 录用日期: 2020年7月22日; 发布日期: 2020年7月29日

摘 要

以银行类指数及成分股的K线收盘价为研究样本, 对银行类指数及其成分股进行了分析。基于银行类指数和其成分股的关系, 首先进行了正回归分析, 并通过回归诊断筛查出银行类指数与各个成分股样本数

据中的异常点。为了使建立的模型更具有实际意义,利用岭估计消除变量间的多重共线性,并基于绝对约束估计和弹性约束估计选择了变量的最优子集,最终给出了符合实际的银行类指数和其成分股之间的线性回归方程。

关键词

聚类分析, 回归分析, 岭估计, 绝对约束估计, 弹性约束估计

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

股票投资是广大投资者投资理财的重要组成部分,银行类企业股票是股票市场中一类重要的股票。银行类股票占有市场的权重较大、一般业绩都比较好、抗风险能力强,它是股民广泛关注的一类股票。因此,有必要研究银行类股票的投资。讨论市场上投资的方法主要有量化投资、基本面分析与技术分析等三种。技术分析是更常用的一类方法,它以股票价格作为主要研究对象,从历史K线图入手,分析总结价格波动规律[1][2]。但少有文章研究银行类指数与其成分股之间的关系。本文以银行类指数及其成分股为研究对象,以源于大智慧应用软件的数据为研究样本,基于R软件,通过聚类分析、回归分析、岭估计等多元统计方法,处理异常数据、消除变量间的多重共线性,得到描述股指与成分股的最优变量,并据此给出复合实际的银行类指数和其成分股之间的线性回归方程。其中样本数据包括银行类指数(指数代码BS991017)及其19支成分股从2014年10月30日至2014年11月18日的15分钟K线收盘价,共计222个样本数据,原始数据从略。在这段时间内,对于缺失的股票数据采用上一时段的数据补齐。

2. 银行类指数成分股的基本统计分析

轮廓图和调和曲线图均为多元数据的可视化方法,它们基于“线”的形式,将高维数据映射到低维空间(三维以下)内表示多元数据出来,同时兼顾信息损失最少和利于肉眼辨识,这有助于进行变量的聚类分析[3]。基于研究样本,由R软件得到其轮廓图与调和曲线图。在样本轮廓图横坐标上取19个点,依次表示19个成分股(即变量),每个拐点代表一支股票;纵坐标对应各个成分股经过标准化变换后的观测值。由轮廓图可以发现6号股票安信信托,指标变动范围非常大。与之相比较,四大国有银行(10号、11号、13号、14号)指标相对较小,而且表现也很平稳。而所研究样本的调和曲线图显示各曲线基本上拧在一起,只是不同成分股的活跃程度大不相同。

变量聚类法是多元统计中的一种重要方法。在系统分析或评估过程中,为避免遗漏某些重要因素,在一开始选取指标时,往往考虑尽可能多的相关因素。这会导致变量过多、变量间的相关度高,给系统分析与建模带来很大的困难[4]。因此,需要研究变量之间的关系,按照变量之间的相似关系将其聚合成若干类,进而找出影响系统的主要因素。现将银行类19支成分股按照2014年10月30号到2014年11月18号的收盘价进行聚类分析。利用SPSS进行成分股聚类,聚类结果显示不管将这19支股票分成几类,陕国投A、安信信托、爱建股份始终单独分成一类,没有和其它16支股票聚类在一起的。同时我们也可以看出南宁银行、宁波银行、北京银行、平安银行也逐渐被分出来,而且单独成为一类。

3. 银行指数和成分股的正回归分析与回归诊断

运用最小二乘法无偏估计模拟参数，对于银行类指数及其成分股，可以得到如下线性回归方程：

$$\hat{y} = -56.38 + 0.93X_1 + 31.76X_2 + \dots + 5.70X_{18} + 32.29X_{19}。$$

其中 \hat{y} 表示银行指数， $X_i (i=1,2,\dots,19)$ 表示 19 支成分股。从回归方程中可以看出，仅截距项是负的。经计算可得此时的残差平方和为 2547.926，与其他估计方法所得的残差相比，该残差平方和是最小的。

根据指数的编制方法可知，回归系数出现负值虽然真实地反应了一段时间内，指数与成分股的实际相关关系，但这种负相关是不正确的，这是因为指数公式中每个成分股对指数的贡献都是正的，即均正相关。此时需要考虑非负最小二乘问题。R 软件中的 `nnls` 数据包提供了非负系数的计算结果，虽然大部分都是正系数，但仍然有些变量的系数为零[5]。由于对于指数公式，所有成分股的权重都不会是零，因此这也是不符合要求的。若成分股对成分指数无贡献，该成分股的存在则不合理了，此时需对模型进行修正从而得到正回归结果。

银行类指数是 19 只成分股股价的加权平均，权重与成分股的市值有关，因此银行类指数是各个成分股的线性函数，它受所有股份股的影响都应该是正的，且不应该有常数项。进行正回归分析，为了避免出现零系数，采用限定最低门限的方法，参考银行类指数板块最小市值股票/总市值 = $4891.9/1142308.8 = 0.004282467$ ，以此为阈值，利用 `nnls` 计算，得到所有系数为正的回归系数。计算出残差平方和为 2566.827，并画出残差图(图 1)和实际走势与正回归预测的效果图(图 2)。从两图可以看出，残差服从白噪声分布，并且预测基本和实际走势吻合。

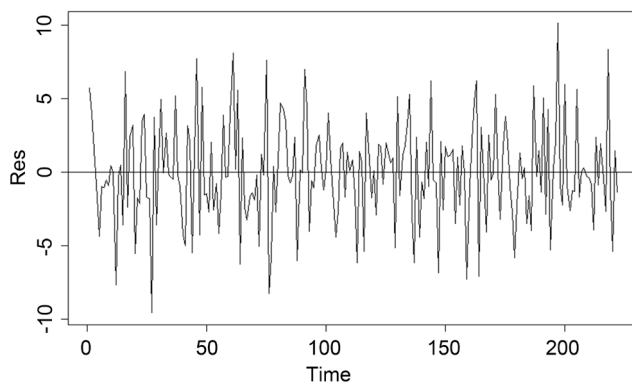


Figure 1. The residual graph of positive regression analysis

图 1. 正回归分析残差图

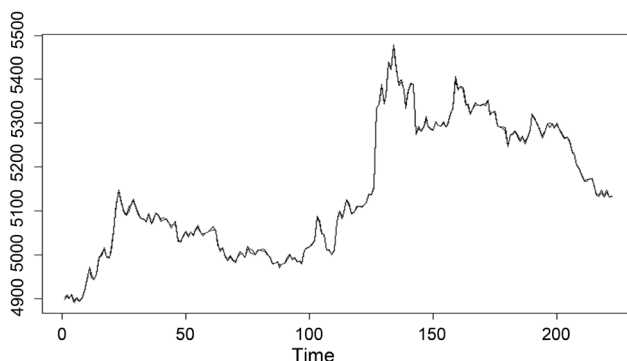


Figure 2. The fitting diagram of positive regression

图 2. 正回归拟合图

为进行残差分析和探测异常点, 需进行回归诊断。首先讨论残差分析, 考虑回归分析模型

$$Y = X\beta + \varepsilon。$$

对于随机误差 ε 需满足如下假设: 1) 正态性; 2) 零均值; 3) 等方差; 4) 独立性。

如果实际数据与这些假设偏离较大, 则最小二乘估计不能保证其仍是无偏估计或有效估计, 也不能进行进一步的假设检验和区间估计, 此时得到的回归方程失去了意义, 由此得到的分析结果不再成立。上述假设不成立便会引起随机误差 ε 分布的变化, 从而导致残差的分布变化。因此可以通过对残差“表现”的分析来诊断关于 ε 的哪些假设不满足, 此即残差分析。这也是回归诊断需要解决的第一个问题[5]。

残差图是进行残差分析较为简单且效果较好的方法。正回归残差图 1 是关于时间序列的残差, 可以看到残差基本上是白噪声的, 从而可推断随机误差满足假设。此外, 也可以绘制残差关于预测值的散点图, 散点图显示残差大致在一个纵坐标为 0 的中心带状区域内随机散落, 无规律性和趋势性, 由此可推断随机误差大致满足四条假设。

另一方面, 检查残差是否服从正态性假定是必要的。关于正态性的假设的检验, 可采用 Shapiro-Wilk 正态性检验(简称 W 检验)作更准确的检验。通过 R 程序计算, $W = 0.9928$, $p\text{-value} = 0.3587$, 即显著的概率是 0.3587, 因此残差满足正态性假设, 接受正态性假设。

异常点的探测是回归诊断的第二个问题。如果一组数据的残差比其它组数据的残差大得多, 我们称其为异常点。由于异常点可能对回归的估计以及其他推断具较大影响, 所以异常点探测尤为必要。我们用基于距离的方法来检测收集的银行类股票数据中的异常点, 主要有 COOK 距离法、WK 距离法以及协方差比准则[5]。可以得到如下检测结果:

1) 用 COOK 距离来诊断, R 程序运行结果显示距离值较大的两个是 135 号($3.170640e-02$)和 27 号($4.866316e-02$), 即 11 月 11 日和 10 月 31 号的数据。经查, 这两天银行类指数出现上涨。同时分析其他板块, 房地产、建筑等指数那两天表现很平庸。数据上确实存在异常。

2) 运用 WK 距离法运行 R 程序, 发现 222 个样本数据中总共 12 个异常值, 由此可以看出, 样本中数据有异常, 但是异常不是太大。

3) 继续用协方差比来诊断, R 程序运行结果显示 132 号样本的协方差比值与 1 的偏离超过 0.92, 对应的是 11 月 11 日的数据, 经查, 这天指数出现上涨。仍然和建筑、房地产等其他板块相比较, 而这几个板块指数表现平稳。这说明, 该样本点对模型的影响还是存在的。

由回归诊断函数可以看出, 不管是用上面三个准则中的哪一个, 均发现 27 号, 132 号, 135 号, 即 10 月 31 日和 11 月 11 日的数据比较异常, 尤其是 11 月 11 日这天的数据。通过观察 11 月 11 日这一天银行类的股票走势可以看出, 这一天的股票是持续涨高的, 这可能是“双 11”掀起的购物狂潮而引起的, 其中中国银行、农业银行, 工商银行涨势更高。

4. 银行指数和成分股回归分析的有偏估计

在回归分析过程中, 进一步分析样本数据可以发现变量间存在严重的多重共线性。经计算, 19 个变量间相关矩阵的条件数为 2885.554, 大于 1000, 最小特征值是 0.0049, 均方误差 MSE 近乎无穷大, 这导致最小二乘无偏估计没有意义, 因而不可再用。有偏估计是解决多重共线性最直接的方法, 比较有代表性的有偏估计有均匀压缩估计、主成分估计、岭估计、刘估计、几乎无偏估计等。

针对前述回归分析中出现的共线性问题, 用岭估计的方法来估计回归系数。岭估计通过解决如下的条件极值问题获得

$$(y - X\beta)'(y - X\beta) + k\beta'\beta,$$

其中 k 拉格朗日乘数。岭估计有如下表达式

$$\hat{\beta}(k) = (X'X + kI)^{-1} X'y.$$

其中 $k \geq 0$ 是岭参数。通过对 k 值的选择,可以减少多重共线性的影响,取不同的 k 值,可以得到不同的估计,因此岭估计 $\hat{\beta}(k)$ 是一个估计类。当 $k = 0$ 时, $\hat{\beta}(0) = (X'X)^{-1} X'y$ 即为常用的最小二乘估计[5][6][7][8]。

由于岭估计含有一个参数 k ,因此作岭估计之前需先选择合适的 k 。实际中如何选择 k ,可有多种方法。首先考虑通过岭迹法选择岭参数。岭迹法的方法是将 p 个分量 $\hat{\beta}_i(k)$ 的岭迹画在同一个图上,参数 k 的选择要使得岭估计 $\hat{\beta}_i(k)$ 的各分量的值大体稳定,并且要兼顾到系数没有不合理的符号,残差平方和上升不太多等[5]。绘出岭迹图,由岭迹图可知当 $k = 0.3$ 时,各条曲线基本趋于稳定。

此外,R 软件还通过函数 select()提供了三种选择岭参数的方法,分别是 HKB (Hoerl-Kennard-Baldwin) 选择, L-W (Lawless-Wang 公式)选择和 GCV (Generalized Cross Validation)最小值选择。

通过 R 程序, HKB 选择、L-W 选择与 GCV 选择分别给出的 k 值为 0.06、0.01 与 0.16。由于岭估计是一种有偏估计,是求解回归系数时的一种无奈之举,因此总是倾向于选择偏差较小的有偏估计。对于岭估计而言, $k = 0$ 就是无偏估计,因此通常选择较小的 k 值。故此时取 L-W 选择给出的参数 $k = 0.01$,因此得到如下经验回归方程

$$\hat{y} = -57.31 + 0.93X_1 + 31.85X_2 + \dots + 5.61X_{18} + 32.36X_{19}.$$

由于原来的最小特征值 0.0049 经过岭参数的附加,变成了 0.0149,均方误差大幅降低,模型中存在的多重共线性明显得到改善,而此时模型的残差平方和为 2548.012,和最小二乘无偏估计相比几乎没有增加。因此岭估计在银行类指数及其成分股的参数模拟上是比较理想的。

5. 银行指数成分股最优子集的选择

对于大量金融数据,不同的变量有着不同的影响,部分变量影响甚微,部分变量具有高度的关联性。考虑到数据收集的成本和分析的时效,并不需要收集全部变量。因此在对金融数据的统计分析中,变量的选择变得尤为重要。此外,基金管理者,投资所有股票并不现实,往往只能选择部分成分股进行投资,用尽量少的变量达到对指数的追踪,从而实现持有的股票组合与股指期货空单形成完全对冲,达到保值的目的[5]。从上面的绝对约束估计可以看出,用该方法并不能很好的实现符合该目的变量最优子集的选择,所以需要进一步进行变量选择。变量选择的方法很多,本文中采用绝对约束估计(Lasso)和弹性约束估计来完成银行类指数成分股最优子集的选择。

绝对约束估计就是要找到满足条件 $\sum_{j=1}^p |\beta_j| \leq t$, 使 $\sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2$ 达到最小的 $\beta_j, j = 1, 2, \dots, p$ 。其中 $t \geq 0$ 是一个调整参数,它控制着估计压缩的程度。设 $\hat{\beta}_j$ 是普通最小二乘估计, $t_0 = \sum |\hat{\beta}_j|$, 当 $t \leq t_0$ 将引发趋向于 0 的压缩,因此绝对约束估计也是压缩估计[9][10][11]。

R 软件的 Lars 软件包提供了解决实际中的变量选择问题的方法。Lars 算法中有一个正则化参数 $s = t / \sum |\hat{\beta}_j|$, 可以采用 BIC (bayes information criterion)准则来选择最佳的 s 。其中 BIC 准则是一种变量选择的准则,即选择使得 BIC 统计量 $BIC = \ln S_E^2 + p \ln(n)/n$ (p 表示所选模型变量的个数(含常数项))达到最小的模型。变量选择的具体步骤是通过 Lars 中的 predicts 函数参数 s 的设置,给出不同的变量子集和回归系数,通过对结果的分析,算出 BIC 值,选择 BIC 达到最小的 s 值即可[5]。求得样本数据 BIC 值最

小值为 10.21, 对应的 $s = 14.19$, 最优子集含有 14 个变量, 残差平方和为 3097.37。由此得到的含有 14 个变量的回归模型为

$$y = -22.66 + 55.29X_2 + 39.89X_3 + \cdots + 9.43X_{17} + 45.93X_{19}。$$

模型中不包含 X_1 (陕国投 A)、 X_5 (安信信托)、 X_6 (爱建股份)、 X_{16} (平安银行)、 X_{18} (宁波银行) 这 5 支股票变量。

进一步, 可以得到自变量和预测值的拟合图(图 3), 从图中可以看出, 拟合效果较好。

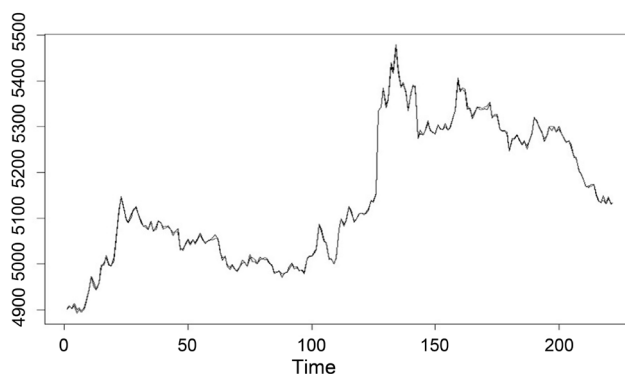


Figure 3. The fitting diagram of absolute constraint estimation
图 3. 绝对约束估计拟合图

尽管在很多情况下绝对约束估计都得到很大认可, 但在某些条件下其有效性也会受到限制, 尤其是在面临高维数据的时候。当变量个数 p 大于样本容量 n 时, 绝对约束估计只能选出 n 个变量, 此时绝对约束估计就不再有效。即使是一般情形 $p < n$ 的情况下, 有时岭回归估计比绝对约束估计表现的要好。

2005 年邹(Zou)和赫斯第(Hastie)提出合并考虑岭回归和 Lasso 的约束方式, 提出了弹性约束估计[12]。其定义如下

$$\tilde{\beta} = \arg \min_{\beta} \left(\sum_i^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right),$$

这等价于找到满足条件 $(1-\lambda) \sum_{j=1}^p |\beta_j| + \lambda \sum_{j=1}^p \beta_j^2 \leq t$, 使 $\sum_i^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2$ 达到最小的 $\beta_j, j=1, 2, \dots, p$

[13]。

显然, 弹性约束估计同时具有绝对约束估计和岭估计的特点。在弹性约束估计中, 当参数 $\lambda = 1$ 时弹性约束估计为岭回归, 当 $\lambda = 0$ 时弹性约束估计为绝对约束估计。

在 R 软件中, 较为方便的是可利用函数 `cv.glmnet` 进行自动 CV 交叉验证, 从而确定出最佳的 λ 值。基于银行类指数及其成分股的样本数据, 得到 $\lambda_{\min} = 2.39$ 。按此参数值, R 程序运行结果保留变量个数是 13 个, 残差平方和为 5233.348。这时相应的得到的最优子集当中仅含有 13 支股票, 其中不包括 X_1 (陕国投 A)、 X_5 (安信信托)、 X_6 (爱建股份)、 X_9 (北京银行)、 X_{16} (平安银行)、 X_{18} (宁波银行) 这 6 支股票。对应的经验回归方程为

$$\hat{y} = 0.84 + 73.79X_2 + 22.65X_3 + \cdots + 14.58X_{17} + 61.45X_{19}$$

预测值拟合图如图 4 所示

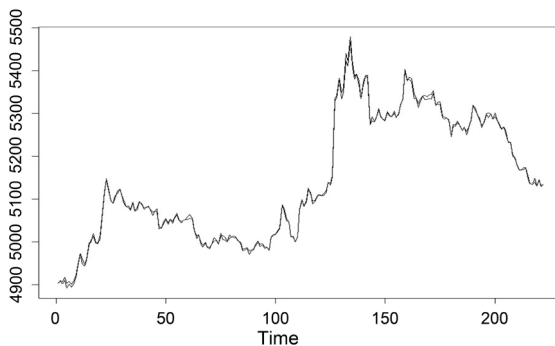


Figure 4. The fitting diagram for an elastic constraint estimate
图 4. 弹性约束估计预测值拟合图

由于 glmnet 程序计算时结果不固定，保留变量个数在 12 和 13 之间变化。通过系数对比，发现 X_7 的系数相对较小($7.480314e-03$)，应该在临界线上，所以导致结果的不确定性。舍弃该变量，最终得到的回归模型中含有 12 个变量

$$\hat{y} = -29.60 + 57.83X_2 + 37.83X_3 + \dots + 7.12X_{17} + 64.71X_{19}。$$

接下来，根据指数的编制方法，进一步进行正回归分析，得到最终的回归模型为：

$$\hat{y} = 56.67X_2 + 37.67X_3 + \dots + 7.59X_{17} + 64.26X_{19}。$$

残差平方和为 3109.354，残差图和预测指数追踪图分别如图 5 与图 6 所示：

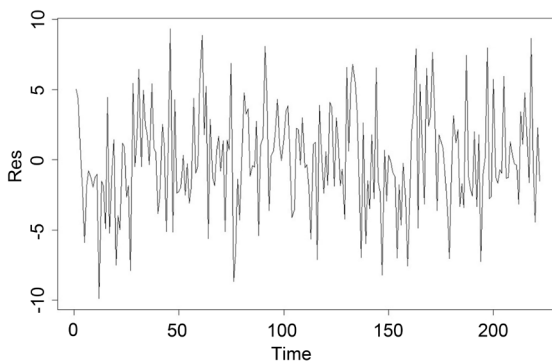


Figure 5. The fitting diagram for an elastic constraint estimate
图 5. 弹性约束估计预测值拟合图

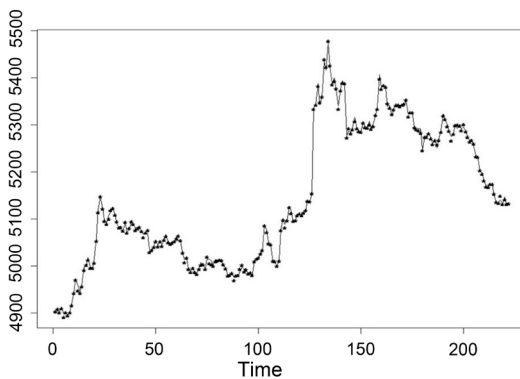


Figure 6. The fitting diagram for an elastic constraint estimate
图 6. 弹性约束估计预测值拟合图

图 6 中符号“*”是实际银行类指数的收盘价，实线是模型预测值。易见，实际走势和 12 只成分股的跟踪走势基本重合，基本跟踪到了指数的运行趋势，与指数运行吻合。

回归模型中各变量对应系数具体如表 1 所示。在最终的回归模型中，不含有的成分股分别为 X_1 (陕西国投 A)、 X_5 (安信信托)、 X_6 (爱建股份)、 X_7 (南京银行)、 X_9 (北京银行)、 X_{16} (平安银行)、 X_{18} (宁波银行)，中国有四大银行回归系数最大，可见其对银行类指数的影响是相对较大，这也与实际相符。

Table 1. The Table of coefficients of positive regression analysis for 12 variables

表 1. 12 个变量正回归分析系数表

X_2 华夏银行	56.67	X_{10} 农业银行	250.02	X_{14} 建设银行	93.30
X_3 民生银行	37.67	X_{11} 工商银行	311.26	X_{15} 交通银行	3.64
X_4 招商银行	30.03	X_{12} 光大银行	73.58	X_{17} 浦发银行	7.59
X_8 兴业银行	37.10	X_{13} 中国银行	301.31	X_{19} 中信银行	64.26

6. 总结

收益与风险是相对应的，也就是两者是相伴而生的。一般地，收益高则风险大，风险小则收益低。正所谓“高风险，高收益；低风险，低收益”。许多人都知道，在有价证券投资中，股票的投资收益高，但统计发现，真正赢利的人一般只占投资者的 1% 左右，而且一旦投资决策错误，在股票价格下跌的情况下会损失惨重。所谓的“投资有风险，入市需谨慎”绝不是危言耸听。当然，股票市场的投资如此，其他项目的投资规律也是这样。

随着资本市场的不断发展，股票数量的增长，基金规模的扩大，信息传导的加速，量化投资策略能够有效规避非理性的负面效应，以客观的方式捕捉市场中的异常讯息，获得超额收益。

量化投资是回避金融风险的有效方法，也是当今非常重要的投资思想，但量化投资的基石是模型，只有不断地根据市场的变化，建立合理实用的投资模型才能有效的实现利益的最大化和有效的规避风险 [14] [15]。

对非结构化数据分析，将看起来相关性并不强的因子纳入量化投资模型，并分配权重，丰富量化投资模型交易的精度和维度。

在本文的分析中，从聚类结果和回归分析过程中，陕西国投 A、爱建股份、安信信托这三支股票的表现都不同于其他股票。这主要是因为这三支股票是普通股，而其他 16 支股票是沪深 300 指数成分股。其中，沪深 300 成分股在运行中有较好的盈利性、成长性和分红收益能力，同时，相对于市场平均水平，其估值优势也比较明显，已经逐渐成为机构投资者乃至整个市场的投资取向标杆。与之相比，作为普通股的这三支股票就会被频繁的“区别”出来。从正回归系数表 1 中可以看出，中国银行、农业银行、工商银行、建设银行四大银行所占权重在 19 支成分股当中是最大的。这是与这四大银行代表着中国最雄厚的资金力量十分不开的。

在量化投资规避风险过程中，需要注意的是数据并非绝对安全，也可能存在风险隐患。现代社会是大数据时代，大数据可能存在这样一种现象：投资者被数据包围，无法判断数据的真实有效性，对于存在缺陷的数据进行建模分析，可能导致所得到的结论不能与现实真实市场环境相一致。例如，当统计模型的样本发生了变动，最终可能导致结论不正确而无法适用于交易决策中。所以在量化投资规避风险时要注意数据陷阱的风险隐患。

基金项目

陆军工程大学基础学科培育金项目(KYJBJQZL1922)；陆军工程大学基础学科培育金项目(KYJBJQZL1921)。

参考文献

- [1] 吕凯晨, 闫宏飞, 陈翀. 基于沪深 300 成分股的量化投资策略研究[J]. 广西师范大学学报(自然科学版), 2019, 37(1): 5-16.
- [2] 曾攀. 基于 R/S 分析法的股市走势分析[J]. 数学理论与应用, 2009, 29(1): 15-18.
- [3] 方开泰. 实用多元统计分析[M]. 上海: 华东师范大学出版社, 1989.
- [4] 何晓群. 多元统计分析[M]. 北京: 中国人民大学出版社, 2012.
- [5] 杨虎. 杨玥含. 金融大数据统计方法与实证[M]. 北京: 科学出版社, 2016.
- [6] Liu, K. (1993) A New Class of Biased Estimate in Linear Regression. *Communications in Statistics—Theory and Methods*, **22**, 392-402. <https://doi.org/10.1080/03610929308831027>
- [7] Yang, H. and Chang, X.F. (2010) A New Two-Parameter Estimator in Linear Regression. *Communications in Statistics—Theory and Methods*, **39**, 923-934. <https://doi.org/10.1080/03610920902807911>
- [8] 薛毅, 陈立萍. 统计建模与 R 软件[M]. 北京: 清华大学出版社, 2007.
- [9] Luo, R., Tsai, C.L. and Wang, H. (2008) On Mixture Regression Shrinkage and Selection via the MR-Lasso. *International Journal of Pure and Applied Mathematics*, **46**, 403-414.
- [10] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [11] Tibshirani, R. (2011) Regression shrinkage and selection via the Lasso: A Retrospective. *Journal of the Royal Statistical Society: Series B*, **73**, 273-282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- [12] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B*, **67**, 301-320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [13] Shen, Y., Han, B. and Braverman, E. (2015) Stability of the Elastic Net Estimator. *Journal of Complexity*, **32**, 20-39. <https://doi.org/10.1016/j.jco.2015.07.002>
- [14] 周浩, 唐琦, 杨雪. 基于大数据的股票多因子量化投资策略优化研究[J]. 商讯, 2019(24): 86-87.
- [15] 贺保国. 资本市场的量化投资策略和风控措施[J]. 中国商论, 2019(8): 51-52.